

OM of: US-09-303-518d-649 to: SwissProt_40:* out_format: pfs
 Date: Jun 30, 2002 8:44 AM
 About: Results were produced by the Gencore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+2p_model -DEV=x11
 -O=/gen2.1/USPPO.spool/US09303518/runat.28062002_142714.4362/app_query.fasta.1.23501
 -DB=SwissProt_40 -OEMT=fasten -SUPPLX=rs -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -TRANS=human40.cdi -LIST=100 -DOCALLIGN=200 -THR.SCORE=pcr
 -THR.MAX=100 -THR.MIN=0 -ALIGN=45 -MODE=LOCAL -OUTPRM=pfis
 -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
 -USER=US09303518 -CGN1.1.440 -NCPU=6 -ICPU=3 -LONGLOC
 -DEV=TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLIPX -WAIT -THREADS=1

Search information block:
 Query: US-09-303-518d-649
 Query length: 4374
 Database: SwissProt_40:*
 Database sequences: 105224
 Database length: 38719550
 Search time (sec): 217.960000

Score List:

| Sequence | Strd | Orig | Zscore | EScore | Len | Documentation |
|--------------------------|------|---------|---------|---------|------|-------------------------------|
| SwissProt_40:ICG1_HAEN + | | 1608.50 | 1426.49 | 6.0e-73 | 1541 | P42782 haemophilus inflen |
| SwissProt_40:ICG1_HAEN + | | 1579.50 | 1400.62 | 1.6e-71 | 1545 | P43385 haemophilus inflen |
| SwissProt_40:ICG1_HAEN + | | 1571.00 | 1393.11 | 4.4e-71 | 1532 | P09790 neisseria gonorrhoea |
| SwissProt_40:ICG1_HAEN + | | 1568.00 | 1389.59 | 6.2e-71 | 1533 | P44969 haemophilus inflen |
| SwissProt_40:ICG1_HAEN + | | 1488.50 | 1317.97 | 7.8e-71 | 1702 | P45384 haemophilus inflen |
| SwissProt_40:ICG1_HAEN + | | 364.50 | 317.42 | 3.5e-11 | 1569 | P52143 escherichia coli |
| SwissProt_40:ICG1_HAEN + | | 295.50 | 260.50 | 8.9e-08 | 911 | P03033 bordetella bronchise |
| SwissProt_40:ICG1_HAEN + | | 288.00 | 259.96 | 9.5e-08 | 912 | P43328 bordetella pertusse |
| SwissProt_40:ICG1_HAEN + | | 288.00 | 253.83 | 2.1e-07 | 910 | P12833 escherichia coli |
| SwissProt_40:ICG1_HAEN + | | 282.00 | 245.80 | 4.3e-07 | 1250 | P45308 escherichia coli |
| SwissProt_40:ICG1_HAEN + | | 275.00 | 239.32 | 9.6e-07 | 1286 | P03155 escherichia coli |
| SwissProt_40:ICG1_HAEN + | | 265.50 | 234.22 | 2.7e-06 | 863 | P33924 escherichia coli |
| SwissProt_40:ICG1_HAEN + | | 252.00 | 215.08 | 1.4e-05 | 1953 | P35666 escherichia coli |
| SwissProt_40:ICG1_HAEN + | | 246.50 | 210.39 | 2.6e-05 | 1953 | P35927 salmonella typhimur |
| SwissProt_40:ICG1_HAEN + | | 242.50 | 208.67 | 3.8e-05 | 1655 | P09K53 r outer membrane pr |
| SwissProt_40:ICG1_HAEN + | | 242.50 | 205.63 | 4.1e-05 | 1249 | P15921 rickettsia rickettsi |
| SwissProt_40:ICG1_HAEN + | | 240.50 | 208.32 | 5.0e-05 | 1335 | P32021 rickettsia rickettsi |
| SwissProt_40:ICG1_HAEN + | | 233.50 | 204.13 | 0.0001 | 1039 | P39180 escherichia coli |
| SwissProt_40:ICG1_HAEN + | | 233.00 | 199.80 | 0.0001 | 1645 | P39180 escherichia coli |
| SwissProt_40:ICG1_HAEN + | | 232.50 | 200.92 | 0.0001 | 1367 | P08640 saccharomyces cerevi |
| SwissProt_40:ICG1_HAEN + | | 231.50 | 199.37 | 0.0001 | 1643 | P053020 r outer membrane pr |
| SwissProt_40:ICG1_HAEN + | | 231.00 | 198.82 | 0.0001 | 1577 | P16466 proteus mirabilis |
| SwissProt_40:ICG1_HAEN + | | 224.50 | 196.28 | 0.0002 | 2021 | P05267 rickettsia conorii |
| SwissProt_40:ICG1_HAEN + | | 224.50 | 192.18 | 0.0003 | 1654 | P05304 r outer membrane pr |
| SwissProt_40:ICG1_HAEN + | | 217.00 | 178.95 | 0.0008 | 3551 | P12255 bordetella pertussis |
| SwissProt_40:ICG1_HAEN + | | 214.00 | 183.05 | 0.0010 | 1699 | P12255 bordetella pertussis |
| SwissProt_40:ICG1_HAEN + | | 211.00 | 183.77 | 0.0014 | 1077 | P02308 drosophila melanog |
| SwissProt_40:ICG1_HAEN + | | 210.50 | 179.69 | 0.0016 | 1656 | P06653 r outer membrane pr |
| SwissProt_40:ICG1_HAEN + | | 210.00 | 179.49 | 0.0016 | 1608 | P15320 serratia marcescens |
| SwissProt_40:ICG1_HAEN + | | 208.50 | 182.80 | 0.0019 | 928 | P45354 haemophilus inflen |
| SwissProt_40:ICG1_HAEN + | | 208.00 | 174.57 | 0.0021 | 2334 | P07833 bacillus subtilis |
| SwissProt_40:ICG1_HAEN + | | 207.50 | 180.92 | 0.0021 | 1045 | P09489 serratia marcescens |
| SwissProt_40:ICG1_HAEN + | | 206.00 | 186.92 | 0.0023 | 1377 | P21561 haloterrax sp. (strain |
| SwissProt_40:ICG1_HAEN + | | 205.50 | 177.01 | 0.0027 | 1341 | P07067 lactococcus lactis |
| SwissProt_40:ICG1_HAEN + | | 205.00 | 173.62 | 0.0030 | 1300 | P14914 rickettsia rickettsi |
| SwissProt_40:ICG1_HAEN + | | 203.50 | 175.49 | 0.0034 | 1300 | P15914 rickettsia rickettsi |
| SwissProt_40:ICG1_HAEN + | | 201.50 | 173.87 | 0.0043 | 1276 | P09289 chlamydia pneumoniae |
| SwissProt_40:ICG1_HAEN + | | 199.50 | 168.72 | 0.0056 | 1902 | P15233 lactococcus lactis |
| SwissProt_40:ICG1_HAEN + | | 199.50 | 168.72 | 0.0056 | 1902 | P15233 lactococcus lactis |
| SwissProt_40:ICG1_HAEN + | | 195.50 | 170.93 | 0.0083 | 959 | P002629 saccharomyces cerevis |

| | | | | | |
|--------------------------|--------|--------|--------|------|------------------------------|
| SwissProt_40:ICG1_HAEN + | 195.00 | 170.86 | 0.0087 | 917 | P45355 haemophilus influenza |
| SwissProt_40:ICG1_HAEN + | 195.00 | 165.54 | 0.0092 | 1723 | P09281 chlamydia pneumoniae |
| SwissProt_40:ICG1_HAEN + | 194.50 | 167.53 | 0.0095 | 1291 | P48258 helicobacter pylori |
| SwissProt_40:ICG1_HAEN + | 194.50 | 165.67 | 0.0097 | 1609 | P09265 chlamydia pneumoniae |
| SwissProt_40:ICG1_HAEN + | 194.50 | 164.26 | 0.0098 | 1902 | P02470 lactobacillus parac |
| SwissProt_40:ICG1_HAEN + | 194.00 | 163.02 | 0.0105 | 2090 | P35658 homo sapiens (hum |
| SwissProt_40:ICG1_HAEN + | 193.00 | 169.07 | 0.0110 | 918 | P34487 caenorhabditis eleg |
| SwissProt_40:ICG1_HAEN + | 192.50 | 165.01 | 0.0149 | 1140 | P06893 saccharomyces cere |
| SwissProt_40:ICG1_HAEN + | 188.50 | 160.29 | 0.0193 | 1616 | P35824 bacillus circulans |
| SwissProt_40:ICG1_HAEN + | 187.50 | 160.25 | 0.0214 | 1460 | P09731 chlamydia muridar |
| SwissProt_40:ICG1_HAEN + | 187.00 | 161.97 | 0.0222 | 1130 | P09731 chlamydia muridar |
| SwissProt_40:ICG1_HAEN + | 187.00 | 159.96 | 0.0226 | 1433 | P45754 bacillus thuringie |
| SwissProt_40:ICG1_HAEN + | 187.00 | 159.69 | 0.0227 | 1480 | P35252 saccharomyces cere |
| SwissProt_40:ICG1_HAEN + | 186.50 | 160.37 | 0.0238 | 1296 | P48247 helicobacter pylori |
| SwissProt_40:ICG1_HAEN + | 186.00 | 160.18 | 0.0251 | 1256 | P52081 staphylococcus aure |
| SwissProt_40:ICG1_HAEN + | 186.00 | 153.29 | 0.0269 | 2843 | P25054 homo sapiens (huma |
| SwissProt_40:ICG1_HAEN + | 185.50 | 161.45 | 0.0261 | 1025 | P09146 chlamydia muridar |
| SwissProt_40:ICG1_HAEN + | 185.50 | 156.57 | 0.0274 | 1829 | P09146 chlamydia muridar |
| SwissProt_40:ICG1_HAEN + | 185.00 | 160.31 | 0.0279 | 1113 | P002630 saccharomyces cere |
| SwissProt_40:ICG1_HAEN + | 185.00 | 158.96 | 0.0282 | 1306 | P23234 saccharomyces cere |
| SwissProt_40:ICG1_HAEN + | 184.00 | 151.50 | 0.0338 | 2845 | P06135 mus musculus (mous |
| SwissProt_40:ICG1_HAEN + | 183.50 | 159.60 | 0.0329 | 1034 | P38008 chlamydia trachom |
| SwissProt_40:ICG1_HAEN + | 183.50 | 159.73 | 0.0335 | 1290 | P55981 helicobacter pylori |
| SwissProt_40:ICG1_HAEN + | 183.50 | 155.54 | 0.0342 | 1672 | P09732 chlamydia muridar |
| SwissProt_40:ICG1_HAEN + | 183.50 | 151.07 | 0.0357 | 2842 | P70478 rattus norvegicus (r |
| SwissProt_40:ICG1_HAEN + | 182.50 | 161.35 | 0.0341 | 797 | P28968 equine herpesvirus |
| SwissProt_40:ICG1_HAEN + | 182.50 | 157.72 | 0.0372 | 1162 | P32323 trypanosoma cruzi |
| SwissProt_40:ICG1_HAEN + | 182.00 | 160.15 | 0.0383 | 827 | P25062 halobacterium volc |
| SwissProt_40:ICG1_HAEN + | 181.50 | 157.10 | 0.0416 | 1125 | P27546 mus musculus (mou |
| SwissProt_40:ICG1_HAEN + | 181.50 | 152.65 | 0.0435 | 1906 | P11799 gallus gallus (ch |
| SwissProt_40:ICG1_HAEN + | 181.00 | 157.88 | 0.0435 | 973 | P09896 chlamydia pneumoniae |
| SwissProt_40:ICG1_HAEN + | 180.50 | 157.32 | 0.0462 | 986 | P44747 escherichia coli |
| SwissProt_40:ICG1_HAEN + | 180.00 | 155.80 | 0.0531 | 1120 | P76072 escherichia coli |
| SwissProt_40:ICG1_HAEN + | 179.50 | 153.86 | 0.0541 | 1337 | P35653 streptococcus do |
| SwissProt_40:ICG1_HAEN + | 179.00 | 155.76 | 0.0549 | 1013 | P08478 chlamydia trachom |
| SwissProt_40:ICG1_HAEN + | 179.00 | 144.77 | 0.0611 | 3726 | P20659 drosophila melanog |
| SwissProt_40:ICG1_HAEN + | 178.00 | 150.74 | 0.0644 | 1754 | P08418 chlamydia trachom |
| SwissProt_40:ICG1_HAEN + | 177.50 | 152.70 | 0.0663 | 1243 | P08418 chlamydia trachom |
| SwissProt_40:ICG1_HAEN + | 176.50 | 151.60 | 0.0745 | 1273 | P38968 saccharomyces cere |
| SwissProt_40:ICG1_HAEN + | 175.00 | 152.15 | 0.0910 | 2333 | P02834 canis familiaris |
| SwissProt_40:ICG1_HAEN + | 174.50 | 152.70 | 0.0910 | 905 | P44602 haemophilus influe |
| SwissProt_40:ICG1_HAEN + | 174.50 | 149.58 | 0.0938 | 1310 | P48253 helicobacter pylori |
| SwissProt_40:ICG1_HAEN + | 174.00 | 149.28 | 0.0975 | 1045 | P29805 serratia marcescens |
| SwissProt_40:ICG1_HAEN + | 173.00 | 145.78 | 0.1141 | 1288 | P09265 chlamydia pneumoniae |
| SwissProt_40:ICG1_HAEN + | 173.00 | 143.04 | 0.1172 | 2426 | P18583 homo sapiens (huma |
| SwissProt_40:ICG1_HAEN + | 172.50 | 145.26 | 0.1209 | 1770 | P084419 chlamydia trachom |
| SwissProt_40:ICG1_HAEN + | 172.00 | 148.10 | 0.1240 | 1199 | P52591 rattus norvegicus (r |
| SwissProt_40:ICG1_HAEN + | 172.00 | 147.05 | 0.1269 | 1592 | P27470 streptococcus do |
| SwissProt_40:ICG1_HAEN + | 171.50 | 149.31 | 0.1320 | 1287 | P48245 helicobacter pylori |
| SwissProt_40:ICG1_HAEN + | 171.00 | 149.31 | 0.1361 | 934 | P43261 escherichia coli |
| SwissProt_40:ICG1_HAEN + | 170.50 | 146.97 | 0.1468 | 1169 | P36170 saccharomyces cere |
| SwissProt_40:ICG1_HAEN + | 169.50 | 147.18 | 0.1629 | 1026 | P03734 bacteriophage t4 |
| SwissProt_40:ICG1_HAEN + | 169.50 | 146.42 | 0.1641 | 1123 | P00520 mus musculus (mou |
| SwissProt_40:ICG1_HAEN + | 168.50 | 147.07 | 0.1812 | 936 | P007591 citrobacter freund |

seq_name: SwissProt_40:ICG1_HAEN
 seq_documentation_block:
 ID ICG1_HAEN STANDARD; PRT: 1541 AA.
 AC P42782;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (ICG1 protease).
 GN IGA.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
 OC Haemophilus.
 NX NCBI_TaxId=727;


```
994 TTTAAGCAGCATATAATGGCAGCAAAAAATCAATGCCAATGACA 1043
    ::::: |||::: |||::: |||::: |||:::
334 LeuasnLysaspSerAlaGlySer...LeuIleGlySerLysThrAsp 349
1044 CAATCTCTGCTAATAGATTAAAAACGACGCTT.....C 1081
    ::::: |||::: |||::: |||::: |||:::
349 rSerThrSerSerAsnGly...LysThrSerThrIleThrGlyGlu 365
1082 AATTGTTAATGTTCTTTATCCGAGCAGCAGCAAGACCTGTTTCAT 1131
    ::: |||::: |||::: |||::: |||::: |||:::
365 ySerLeuasnValAspLeuAlaAspGlyLysAspPro..... 378
1132 GCTGCAAGGTGCTGTCACAGCTTATCGACCCAGACTGAATATAGAGAAA 1181
    |||::: |||::: |||::: |||:::
379 .....AsnHisGlyLysSe 383
1182 TATTCTTTATTTAGCAAGGAAAAAGCCAAATTGATTAATCCAGCAACA 1231
    ::::: |||::: |||::: |||::: |||::: |||:::
383 rValThrPhe.....GluGlySerGlyThrLeuThrLeuAsnAsn 398
1232 TCAATCAAGGTGCTGAGAGATTATTTCCAGAGAGATTTTAAGGCTC... 1278
    |||::: |||::: |||::: |||::: |||::: |||:::
398 lAspGlnGlyAlaGlyLeuPhePheGlyAspTyrGluValLys 414
1279 TCGCTGCAAAATACGAAACTTGGCAAGGCGCGCTTCATATCAGTGA 1328
    ::::: |||::: |||::: |||::: |||::: |||:::
415 GlyThrSerAspAsnThrThrTyrLysGlyAlaGlyValIleGlu 431
1329 AGCAGTACCGTTACTTGGAAAGTAAAGCGCGTGGCAAGCGCCCTGT 1378
    |||::: |||::: |||::: |||::: |||::: |||:::
431 uGlyLysThrValThrTyrLysValHisAsnProGlnTyrAspArgLeu 448
1379 CCAAAATCGGCAAGGCGACGCTGACGTTCAAGCCAAAGGGGAAACCAA 1428
    ::::: |||::: |||::: |||::: |||::: |||:::
448 lAlaLysIleGlyLysGlyThrLeuIleValGluGlyThrGlyAsp 464
1429 GCGTGCATGACGGTGGCGAGCGTACAGTATTTGGATCAGCAGGCA 1478
    |||::: |||::: |||::: |||::: |||::: |||:::
465 GlySerLeuLysValGlyAspGlyThrValIleLeuLysGlnGlnThr 481
1479 CGATAAAGGCAAAACAGGCTTATGTAATCGGCTTGGTGCAGCGCA 1528
    ::::: |||::: |||::: |||::: |||::: |||:::
481 nGlySerGly...GlnHisAlaPheAlaSerValGlyLysGly 497
1529 GGGTACGCTGCACTGAATGCCGATATACGTTACCCGACAAATCG 1578
    |||::: |||::: |||::: |||::: |||::: |||:::
497 rGSetThrLeuValLeuAsnAspAspLysGlnValAspProAsnSer 513
1579 TATTGCGCTTTCGCGCGGACGTTGGATTAAACGGGCATTTGCTTTC 1628
    |||::: |||::: |||::: |||::: |||::: |||:::
514 TyrPheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnSerLeu 530
1629 GTTCACCGCATTTCAAAATACCGATGAAGGGCGATGTTGTCACACCA 1678
    |||::: |||::: |||::: |||::: |||::: |||:::
530 rPheAspHisIleLeuArgAsnIleAspAspGlyAlaArgLeuValAsn 547
1679 ATCAAGACAAAGAAATCCACGCTTACCATACAGC..... 1713
    |||::: |||::: |||::: |||::: |||::: |||:::
547 smMetThrAsnAlaSerAsnIleThrIleThrGlyLysSerLeuIle 563
1713 ..... 1713
564 AspProAsnThrIleThrProTyrAsnIleAspAlaProAspGluAsp 580
1713 ..... 1713
580 nProTyrAlaPheArgArgIleLysAspGlyGlnLeuTyrLeuAsn 597
1713 ..... 1713
597 euGluAsnTyrThrTyrTyrAlaLeuArgLysGlyAlaSerThrArg 613
1713 ..... 1713

614 GluLeuProLysAsnSerGlyGluSerAsnGluAsnThrLeuTyrMet 630
1714 ....AATAAGATATTGCTACACCGGCAATTAACACAGCTTGGATGCA 1759
    ::::: |||::: |||::: |||::: |||::: |||:::
630 yLysThrSerAspGluAlaLysArgAsnValMetAsnHisIleAsn 647
1760 AAAAAAAATTTGCTCAACAGGTTGGTTGGCGAAGAAATACACACCA 1809
    |||::: |||::: |||::: |||::: |||::: |||:::
647 lArgMetAsnGlyPheAsnGlyTyrPheGlyGluGln...GlyLys 662
1810 ACGAAGCGGCGCTCAACCTGTTTACCAAGCCCGCCGACAGAACCCGAC 1859
    ::::: |||::: |||::: |||::: |||::: |||:::
663 AsnAsnGlyAsnLeuAsnValThrPheLysGlyLysSerGlnAsn 679
1860 CTTGCTGCTTTCGCGGACAAATTTAAAGGCAATCAACGCAACCA 1909
    ::::: |||::: |||::: |||::: |||::: |||:::
679 gPheLeuLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuThrValGlu 696
1910 ACGCAAACTGTTTTCAGCGGACAGCAACACCGCAGCTTACATCAT 1959
    ::::: |||::: |||::: |||::: |||::: |||:::
696 ySgLyThrLeuPheLysSerGlyArgProThrProHisAlaArgAsp 712
1960 TTTAAAGCAGCATTTGTCGCAAAAAGAGGC.....ATTCTCGCGGGA 2003
    ::::: |||::: |||::: |||::: |||::: |||:::
713 AlaGlyIleSerSerThrLysLysAspProHisPheAlaGluAsn 729
2004 AATCGTGTGGACACAGCTGATCAACCGCATTTAAAGCGGAAACT 2053
    |||::: |||::: |||::: |||::: |||::: |||:::
729 uValValValGlnLysAspArgPheAsnArgAsnPheLysAlaThr 746
2054 TCCAAATTTAAAGCGGCGACGCGCTGCTTTC...CGCAATGTTGCCAA 2100
    ::::: |||::: |||::: |||::: |||::: |||:::
746 eLAsnValThrGlyAsnAlaSerLeuTyrSerGlyArgAsnValAla 762
2101 GTGAAGGCGCATTTGCAATTCAGCAATCAGCCCAACATTTTGGTGT 2150
    |||::: |||::: |||::: |||::: |||::: |||:::
763 lIleThrSerAsnIleThrAlaSerAsnLysAlaGlnValHisIleGly 779
2151 CGCAGCGCATCAAGCAGCAACATCTGTACAGCTTGGATGAGCGGTC 2200
    ::::: |||::: |||::: |||::: |||::: |||:::
779 r.....LysThrGlyAspThrValLysValArgSerAspTyrThr 794
2201 TGACAAATTTGTGCGAAAAAACCATTAACGACGATAAGATGCTTCA 2250
    ::::: |||::: |||::: |||::: |||::: |||:::
794 yValThrCysThrThrAspLysLeuSer...AspLysAlaLeuAsn 809
2251 TTGACTAAGACGACATCAGCGGCAATGTGATCTTGCAGTACGCTCA 2300
    ::::: |||::: |||::: |||::: |||::: |||:::
810 PheAsnProThrAsnLeuArgLysAsnValAsnLeuThrGluSerAla 826
2301 TTTAAATCTCAGAGGCTTCCACTCAACGCGCAATTTAGTGAATG 2350
    ::::: |||::: |||::: |||::: |||::: |||:::
826 nPheValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSerArg 842
2351 GCGATACACCTTATACAGTCAGCAGCAAGCGCCACCAAAAGGCAACT 2400
    |||::: |||::: |||::: |||::: |||::: |||:::
842 LysAsnSerGlnValArgLeuThrGluAsn..... 851
2401 AGCCTGTGGCAATGCCAAGCAGCATTTAATCAGCAGCATTTAAAG 2450
851 ..... 851
2451 CAACACATCGGCTTGGGCAATGCTTATTATTAACGACGACGCG 2500
851 ..... 851
2501 TACAAAGCGAGCTGTGACGTTTCGCGCAAGCTAAGCAACGTAAGC 2550
851 ..... 851
2551 CATTCGCACTCAAGGTATGTCTCCAGCGATTAAGCAGTATTTCCA 2600
```

```
851 ..... 851
2601 TTTTGAAGCAGCGCGTTACCGGACAAATCAGCGCGGCAAGATACGG 2650
851 ..... 851
2651 CATTACACTTAAAAAGCAGCGAATGACGCTCCGTCAGGACGGAATTA 2700
852 ..... 852
2701 GGCAATTTAAACCTTGACAGCGCACCATTTACATCAATTCGCGCTATCG 2750
863 HisGlnLeuAspLeuAlaAsnGlnHisIleHisLeuAsnSerAlaAsp 879
2751 CCAAGATGCGGCAAGGCGCAACCGGCACTGCGACAGATGCGCGCGCC 2800
879 nSerAsn..... 881
2801 GCCGTTGCGCGCGTTCGCGCGCTTATTCGTTACCGCGCACT 2850
882 ..... 884
2851 TCGGTAGATCCGTTTCAACAGCGTGCAGTAAACGCGCAATTTGAACGG 2900
885 ..... 896
2901 TCAGGACATTCGCGCTTTATGTCGGAATCTTCGCTACCGGACGACA 2950
896 yAsnGlySerPheTyrLeuThrAspLeuSerAsnLysGlnGlyAsp 913
2951 AATTGAAGCGGCGGAAAGTTCCGAGGCGACATTACACCTTGCGGCTCAAC 3000
913 yValValValThrLysSerAlaThrGlnLysPheThrLeuGlnValAla 929
3001 AATACCGGCAACGACCTGCAGCGCTGCAACATTTGACGTAAGTGAAGG 3050
930 AspLysThrGlnGlyLupro...AsnHisAsnGlnLeuThrLeuPheAsp 945
3051 AAAAGACACAAACCGCTGTCGGAACCTTAATTCAACCTGCAAAACG 3100
945 a.....SerLysAlaGlnArgAspHisLeuAsnValSerLeuValGly 960
3101 AACAGCTGATGCGCGCGCTGCGCTTACCAATCATCCGCAAGACGCG 3150
960 snThrValAspLeuGlnValAlaTrpLysTyrLysLeuArgAsnValAsnGly 976
3151 GAGTTCCCGCTGCATATCCGCTCAAGACAA..... 3183
977 ArgTyrAspLeuTyrAsnProGlnValGlnLysArgAsnGlnThrValAs 993
3183 ..... 3183
993 pThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspValProSerV 1010
3184 ..... 3207
1010 alProSerAsnAsnGlnGlnIleAlaArgValAspGlnAlaProValPro 1026
3208 ..... 3215
1027 ProProAlaProAlaThrProSerGlnThrThrGlnThrValAlaGlnAs 1043
3216 CAAAAGACAGCGGCAAAAA.....GACAAACCGCAAAAGCCTTGACG 3256
1043 nSerLysGlnGlnSerLysThrValGlnLysAsnGlnGlnAspAlaThrG 1060
3257 CGCTGATTGCGCGCGCGCGCATGCGCGTGA..... 3288
1060 lThrThrAlaGlnAsnArgGlnValAlaLysGlnAlaLysSerAsnVal 1076
3289 AAGACAGAAAGCGTTGCCGACCGCGCGGACGAGCGAGCGGAGAAATGT 3338
1077 LysAlaAsnThrGlnThrAsnGlnValAlaGlnSerGlySerGlnThrLys 1093

3339 CGGCAATTATGACGCGGAG.....GAG 3361
1093 sGlnThrGlnThrThrGlnThrLysGlnThrAlaThrValGlnLysGln 1110
3362 AGAAAAAGCGGTCAGCGCGGATTAAGACACCGCTTGCGGAAA..... 3405
1110 lLysAlaLysValGlnThrGlnLysThrGlnGlnValProLysValThr 1126
3406 ..... 3416
1127 SerGlnValSerProLysGlnGlnGlnSerGlnThrValGlnProGlnAl 1143
3417 GGA..... 3420
1143 aGlnProAlaArgGlnAsnAspProThrValAsnIleLysGlnProGlns 1160
3420 ..... 3420
1160 ergLlnThrAsnThrAlaAspThrGlnGlnProAlaLysGlnThrSer 1176
3420 ..... 3420
1177 SerAsnValGlnGlnProValThrGlnSerThrThrValAsnThrGlnLys 1193
3421 ..... 3445
1193 nSerValValGlnAsnProGlnLysAsnThrThrProAlaThrGln...P 1209
3446 CC..... 3459
1209 rThrValAsnSerGlnSerSerAsnLysProLysAsnThrHisArgArg 1225
3460 GCCCGCGCGGATTTGCCG...CACTGCACCCCAACCGCACCGCCCAAC 3506
1226 SerValArgSerValProHisAsnValGlnProAlaThrThrSerSerAs 1242
3507 GCAGCGC..... 3538
1242 nAspArgSerThrValAlaLeuLysAspLeuThrSerThrAsnThrAsnA 1259
3539 CGCGTTGAGTGAATTTCCGCC.....ACGCTCAACAGC 3573
1259 lValLleuSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAsnVal 1275
3574 GTTTCGCGGTACAGGACGAATTTAGCGCGTA...TTGCCGAAGACCG 3620
1276 GlyLysAlaValSerGlnHisIleSerGlnLeuGlnMetAsnAsnGln 1292
3621 CGGCAACGCGGTTTGACAAAGCGGCATCGGGACACCAACACTACCGTT 3670
1292 yGlnTyrAsnValTyrValSerAsnThrSerMetAsnLysAsnTyrSerS 1309
3671 CGCAAGATTTCCGCGCTACCGCGCAACAAACGACCTGCGCAATCGT 3720
1309 erSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1325
3721 ATGCAGAAAACTCGCGACGCGG...CGCGTCGCACTCGTTTGCGA 3767
1326 TrpAspLlnThrIleSerAsnAsnValGlnLeuGlyValPheThrTyr 1342
3768 CAACCGGACCGAAGAACCTTCAGCAGCGCATCGGCACTCGCGACGCG 3817
1342 rValArgAsnSerAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 1358
3818 TTGCGCACGCGCGCGTTTTCGCAATATGCGCATCGAC...AGGTTCTAC 3864
1358 euAlaGlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTyrTyr 1374
3865 ATGCGCATACGCGCGCGCGGTTTTCAGCAGCGGACGCTTTCAGACGG 3914
1375 LeuGlyLleAspLeuGlyTyrGlyLysPheGlnSerLysLeuGlnThrAs 1391
```



```
720 rSerThrLysLysAspSerHisPheSerIu.....AsnA 732
2000 GGGAAATCGTGTGGACCAACGACTGCATACCGGCACTTAAAGCGGAA 2049
:::||||| :||||| :||||| :||||| :|||||
732 snGluValValIgluAspSPTrIleAsnArgSerPheLysAlaThr 748
2050 AACTTCCAAATTAAGGCGGACGCGGTGTTCC...CGCAAGTTTCG 2096
||||| :||||| :||||| :||||| :|||||
749 AsnIleAsnValThrAsnAsnAlaThrLeuYrSerGlyArgSerValGl 765
2097 CAAGGAAAGGCGATTGGCATTGTGCAATACAGCCCAAGCATTTTTCG 2146
::: :||||| :||||| :||||| :|||||
765 uSerIleThrSerAsnIleThrAlaSerAsnAlaLysValHisIleG 782
2147 GTGTCCACCGCATCAAGCCACACATCTGACAGTTCAGGTCGAGGACG 2196
|| :||||| :||||| :||||| :|||||
782 LysYr.....LysAlaGlyAspThrValCysValArgSerAspYrThr 796
2197 GGTCTGCAAAATGTGTGCAAAAAAACCATTCAGCAGATAAGTATTGC 2246
|| :||||| :||||| :||||| :|||||
797 GlyYrValThrCysThrThrAspLysLeuSer...AspLysAlaLeuAs 812
2247 TTCATTGACTAAGACGACATCAGCGGCAATGTCATCTGCCATTCACG 2296
||||| :||||| :||||| :||||| :|||||
812 uSerPheAsnProThrAsnLeuArgGlyAsnValAsnLeuThrGluSerA 829
2297 CTCATTAAATCTCACAGGCGTTGCCACATCCACGCGCAATCTTAGTCA 2346
|| :||||| :||||| :||||| :|||||
829 laAsnHeValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSer 844
2347 AATGGGATACAGGTTATACAGTCCACCAACAGCCCAACCCAAAAAGCGCA 2396
:::||||| :||||| :||||| :||||| :|||||
845 ArgGlyAsnSerGlnValArgLeuThrIuAsn..... 855
2397 CCTTAGCCTCGTGGCAATGCCCAAGCAACATTAATCAACGACATTA 2446
855 ..... 855
2447 AAGGCAACACATCGGCTTCGGGCAATGCTTATTAACTAAGGACAC 2496
855 ..... 855
2497 GCCGTACAAAAGCGAGTGTGACGCTTCGGGCAACGCTAAGGAAAGCT 2546
855 ..... 855
2547 AAGCCATTCCGCACTCAAGGTAATGTCTCCCTAGCCGATAAGGCACTAT 2596
855 ..... 855
2597 TCCATTTTGAAGACGCCGCTTTACCGGACAAATCAGCGCGCGCAAGAT 2646
855 ..... 855
2647 ACGGATTACACTTAAAGACAGCAATGAGCGTCCGCTCAGGACGGA 2696
||||| :||||| :||||| :||||| :|||||
856 .....SerHisThrIleuThrGlyAsnSerAs 865
2697 ATTAGGCAATTTAACCTTGACAAGCCACCATTAACATTCGCGCT 2746
::: :||||| :||||| :||||| :||||| :|||||
865 pValHisGlnLeuAspLeuAlaAsnGlyHisIleHisLeuAsnSerAla 882
2747 ATCCGACAGATGCGGACGAGGGCGCAAAACGCGAGTCGACAGATGCGCGG 2796
::: :||||| :||||| :||||| :||||| :|||||
882 spAsnSerAsn..... 885
2797 CGCCGCCGTTGCGCGCTTCGCCCTTATTCGTTACACGCGC 2846
:::||||| :||||| :||||| :||||| :|||||
886 .....AsnValIhr..... 888
2847 AACTTCGTAGAAATCCGTTTCAACAGCTGACGAGTAAAGCGCAATTTGA 2896
:::||||| :||||| :||||| :||||| :|||||
889 .....LysTyrAsnThrLeuThrValAsn...SerLeuS 899
```

```
2897 ACGGTACAGGCAACATCCGCTTTATGTCCGAACCTTCGTCACCGACG 2946
:::||||| :||||| :||||| :||||| :|||||
899 erGlyAsnGlySerPheYrYrLeuThrAspLeuSerAsnLysGlnGly 915
2947 GACAAATTTGAAGTCCGCGGAAAAGTTCCGAAGCCACTTACACCTTGGCGGT 2996
||||| :||||| :||||| :||||| :|||||
916 AspLysValValValThrLysSerAlaThrGlyAsnPheThrLeuGlnVal 932
2997 CAACATACCGGCAACGACCTGCAAGCCTCGAACAATTCAGCGTAAGG 3046
|| :||||| :||||| :||||| :||||| :|||||
932 laAlaAspLysThrGlyGluPro...AsnHisAsnIleuThrLeuPheA 948
3047 AAGCAAAAGACAAACAACCGCTGCCGAACCTTAATTTACCCCTGCA 3096
::: :||||| :||||| :||||| :||||| :|||||
948 spAla.....SerLysAlaGlnArgAspHisLeuAsnValSerLeuVal 962
3097 AACGACACGTCGATCCGCGCGCTGCGGCTTACCACTCATCCGCAAGA 3146
::: :||||| :||||| :||||| :||||| :|||||
963 GlyAsnThrValAspLeuGlyAlaTrpLysTyrLysLeuArgAsnValAs 979
3147 CCGCGAGTTCGCGCTGCATATCCGTCAAAGAACAA..... 3183
::: :||||| :||||| :||||| :||||| :|||||
979 nGlyArgYrAspLeuYrAsnProGluValGluLysArgAsnGlnThrY 996
3183 ..... 3183
996 AlaSPThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspValPro 1012
3184 .....GAGCTTCCGACAAACTCGGCAAG..... 3207
::: :||||| :||||| :||||| :||||| :|||||
1013 SerValProSerAsnAsnGlnGluIleAlaArgValAspGluAlaProVa 1029
3208 ..... 3208
1029 lProProAlaProAlaThrProSerGluThrThrGluThrValAlaG 1046
3212 AAGCAAAAACAGCGCGGAAAA.....GACAAACGCGCAAAAGCTT 3252
|| :||||| :||||| :||||| :||||| :|||||
1046 luAsnSerLysGlnLeuSerLysThrValGluLysAsnGluAlaSpAla 1062
3253 GACGCGTGAATGCGCGCGCGGATGCGTCGAA..... 3288
|| :||||| :||||| :||||| :||||| :|||||
1063 ThrGluThrThrAlaGlnAsnArgGluValAlaLysGluAlaLysSerAs 1079
3289 ...AAGCAGAAAAGCTTCCGGAACCGCGCGCGGACGAGCGGCGGAAA 3334
||||| :||||| :||||| :||||| :|||||
1079 nValLysAlaAsnThrGlnThrAsnGluValAlaGlnSerGlySerGluT 1096
3335 ATGTCGCGATTATGACAGCGGAG..... 3357
::: :||||| :||||| :||||| :||||| :|||||
1096 hrLysGluThrGlnThrThrGluThrLysGluThrAlaThrValGluLys 1112
3358 GAAAGAAAAACGCGTGACAGCGGATTAAGACAGCGCTTGGCGAAA.. 3405
||||| :||||| :||||| :||||| :|||||
1113 GluGluLysAlaLysValGluThrGluLysThrGlnGluValProLysVa 1129
3406 ..... 3406
1129 lThrSerGlnValSerProLysGlnGluGlnSerGluThrValGlnProG 1146
3413 AAGCGGAA..... 3420
:::||||| :||||| :||||| :||||| :|||||
1146 lnaLagLupProAlaArgGluAsnAspProThrValAsnIleLysGluPro 1162
3420 ..... 3420
1163 GlnSerGlnThrAsnThrThrAlaAspThrGluGlnProAlaLysGluTh 1179
3420 ..... 3420
1179 rSerSerAsnValGluGlnProValThrGluSerThrThrValAsnThrG 1196
```

```

3421 .....ACCGGCGCGGTACACCGCC 3441
1196 LysAsnSerValValGluAsnProGluAsnThrThrProAlaThrThrGln 1212
3442 TTCGCC.....CGGCGCCG 3455
1213 ...ProThrValAsnSerGluSerSerAsnLysProLysAsnArgHisArg 1228
3456 CGCGCGCGCGCGGATTTGCCG...CAATGCAACCCCAACCGCAGCGCC 3502
1228 gATgSerValArgSerValProHisAsnValGluProAlaThrThrSerS 1245
3503 AACCGCAGCGC.....GACCTGATCAGCCGTTATGCC 3534
1245 eRAsnAspArgSerThrValAlaLeuCyAspLeuThrSerThrAsnThr 1261
3535 AATAGCGGTTGGTGAATTTCCGCC.....ACGCTCAA 3569
1262 AsnAlaValLeuSerAspAlaAlaValAlaGlnPheValAlaLeuS 1278
3570 CAGCGTTTTCGCGTACAGCAGCAATTAAGCCGCTA...TTTGCCTGAAG 3616
1278 ValGlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnSng 1295
3617 ACCGCGCGCAACCGCGTTTGACAAAGCGGCATCCGGGACACCAAACTAC 3666
1295 LuGlyGlnTyraAsnValTyraValSerAsnThrSerMetAsnLysAsnTy 1311
3667 CGTTTCGCAAGATTTCCGCGCTACCGCCCAAAACGACCTGGCGCAAT 3716
1312 SerSerSerGlnTyraArgArgPheSerSerLysSerThrGlnThrGln 1328
3717 CGGTATGCAGAAAAAAGCTCGGACGCGG...CGCGTGCAGATCCTGTTT 3763
1328 uGlyTyraPAspGlnThrIleSerAsnAsnValGlnLeuGlyValPheT 1345
3764 CGCAGACCGGACCGGAAACACTTCGACGAGCGCATCGGCACTCGGCA 3813
1345 hTyTyraValArgAsnSerAsnAsnAsnAsnAsnAsnAsnAsnThr 1361
3814 CGGCTTGGCGCGCGCGGCTTTTGGCGCATACGGCATCGAC...AGGTT 3860
1362 ...LeuAlaGlnValAsnPheTySerLysTyTyraLAspAsnHisStr 1377
3861 CTATAGCGGATTCAGCGCGCGCGGCTTTAGACGCGGACGCTTCAG 3910
1377 pTyLeuGlyIleAspLeuGlyTyGlyLysPheGlnSerLysLeuGlnT 1394
3911 ACGGATCGAGGAGCAAAATCCGCGCGCGCTGTGCATTACGCGATTAG 3960
1394 hArgHisAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThr 1410
3961 GCAGCATACCGCGCGGCTTTGGCGGATTCGGCATCGACCGGCAATCGG 4010
1411 AlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValG 1427
4011 CGCAGCGCGGCTATTTCGTCAAAAAGCGATTACCGCTACGAAAAAGTCA 4060
1427 yValArgTySerTyTyraSerAsnAlaAspPheAlaLeuAspGlnAlaA 1444
4061 AATAGCGACCGCGCGCTTCGATCAACCGCTACCGCGCGGCAATTAG 4110
1444 gIleLysValAsnProIleSerValLysThrAlaPheAlaGlnValAsp 1460
4111 GAGATTTATTCATCAACCGCGCGCAACATTCATTCAGCGCTTATT 4160
1461 LeuSerTyThrTyHisLeuGlyLys...PheSerValThrProIleLe 1476
4161 GAGCTGTTCCTATACGATCGCGCTTCGGAAGAGTCCGAACCGCGTCA 4210
1476 uSerAlaAlaGlyTyr...AspAlaAsnGlnGlySerGlyLysIleAsnValA 1492
4211 AATACCGCGCTATTGGCTCAGATTTCCGCAAAACCGCGAGTGGCAATGG 4260

```

```

seq_name: SwissProt_40:IGA_NEIGO
seq_documentation_block:
ID IGA_NEIGO STANDARD; PRT; 1532 AA.
AC P09790;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Iga-specific serine endopeptidase precursor (EC 3.4.21.72) (Iga
DE protease).
DE
GN IGA.
OS Neisseria gonorrhoeae.
OX Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NCBI_TaxID:485;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN=MS11;
RA MEDLINE=87115823; Pubmed=3027577;
RT Pohlner J., Halter R., Beyreuther K., Meyer T.F.;
RT "Gene structure and extracellular secretion of Neisseria gonorrhoeae
RT Iga protease."
RT Nature 325:458-462(1987).
RL [2]
RP ACTIVE SITE.
RX MEDLINE=90154052; Pubmed=2105953;
RA Bachovchin W.W., Plant A.G., Flentke G.R., Lynch M., Kettner C.A.;
RT "Inhibition of Igal proteinases from Neisseria gonorrhoeae and
RT Hemophilus influenzae by peptide prolyl boronic acids."
RL J. Biol. Chem. 265:3738-3743(1990).
CC -1- FUNCTION: THIS PROTEASE IS SPECIFIC FOR IMMUNOGLOBULIN A.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF AUTOPROTEOLYIS.
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYIS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL, X04835; CAN28538.1; -.
DR PIR, A26039; A26039.
DR MEROPS, S06.001; -.
DR InterPro: IP000710; IGA_S6.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; zymogen; Autocatalytic cleavage;
KW Transmembrane; Signal.
FT SIGNAL 1 27
FT CHAIN 28 986 IGA-SPECIFIC SERINE ENDOPEPTIDASE.

```

| FT | PROPEP | 987 | 1532 | HELPER PEPTIDE. |
|----|----------|----------|-------------|--------------------------|
| FT | ACT_SITE | 278 | 278 | POTENTIAL. |
| FT | SITE | 986 | 987 | CLEAVAGE (AUTO-). |
| FT | SITE | 1018 | 1019 | CLEAVAGE (AUTO-). |
| FT | SITE | 1121 | 1122 | CLEAVAGE (AUTO-). |
| SO | SEQUENCE | 1532 AA: | 1669976 MW: | 68FEF4112BD2E240D CRC64: |

```

alignment_scores:
  Quality: 1571.00
  Ratio: 1.791
  Gaps: 46
Percent Similarity: 50.694
Percent Identity: 27.341

```

```
alignment_block:
```

US-09-303-518D-649 X IGA_NEIGO

Align seg 1/1 to: IGA_NEIGO from: 1 to: 1532

```

49  AAAACGGCCCGCAGCCGCTTCCTCCGCTTACCATCATCCGCTGC 98
      |||::: |||::: |||::: |||::: |||::: |||:::
2   LysAlaLysArgPheLysIleAsnAlaIleSerIleuSerIlePheLui 18
      |||::: |||::: |||::: |||::: |||::: |||:::
99  GTTCGGCATTCCTCCCCAACGCTTGGCGGACACACTTATTTCCGATCA 148
      |||::: |||::: |||::: |||::: |||::: |||:::
149  ACTACCAATACTATCGGAGCTTTGGCGAATAAAGCAGCTTTGGCAGTC 198
      |||::: |||::: |||::: |||::: |||::: |||:::
35  spLYrGInIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51
      |||::: |||::: |||::: |||::: |||::: |||:::
199  GGGCGGAAAGATATTTGAGGTTTACACAAAAGAGGAGATTTGGCGGCA 248
      |||::: |||::: |||::: |||::: |||::: |||:::
52  GlyAlaThrAspLeuSerValLysAsnLysArgGlyGluAsnIleGlyAs 68
      |||::: |||::: |||::: |||::: |||::: |||:::
249  ATCATATCACAAAAGCCCGCATGATTGATTTTCTGTGCTGCGCGTAAC. 297
      |||::: |||::: |||::: |||::: |||::: |||:::
68  nAlaLeuSerAsnValPrometIleAspPheSerValAlaAspValAsnI 85
      |||::: |||::: |||::: |||::: |||::: |||:::
298  . GCGCGTGGCGGATTCGTGGCGGATCAATATATTGAGACCGTGGCAT 345
      |||::: |||::: |||::: |||::: |||::: |||:::
85  ySaArgIleAlaThrValValAspProGlnIlyrAlaValSerValLysShs 101
      |||::: |||::: |||::: |||::: |||::: |||:::
346  .....AACGGCGGCTATACACAGCTTGA 368
      |||::: |||::: |||::: |||::: |||::: |||:::
102  AlAlaLysAlaGlyValHisThrPheLyTyrGlyGlnTyrAsnGlyHisAs 118
      |||::: |||::: |||::: |||::: |||::: |||:::
369  TTTTGGCGCGAGAGGAAGAATCCCGATCAACATCGTTTACTTATAAAA 418
      |||::: |||::: |||::: |||::: |||::: |||:::
118  nAspValAlaAspLysGluAsn.....GluTyrArgLy 129
      |||::: |||::: |||::: |||::: |||::: |||:::
419  TTGTGAACGAGATTAATTATTAAGCAGGACGTAAAGCCGATCTTTATGGC 468
      |||::: |||::: |||::: |||::: |||::: |||:::
129  aLValaGluGlnAsnAsnTyrGluPro.....HisLysAlaTrp 141
      |||::: |||::: |||::: |||::: |||::: |||:::
469  GGC.....GATTATCATATATCCCGGTTTGGCA 494
      |||::: |||::: |||::: |||::: |||::: |||:::
142  GlyAlaSerAsnLeuGlyArgLeuGlnAspTyrAsnMetalAlaArgPheAs 158
      |||::: |||::: |||::: |||::: |||::: |||:::
495  TAAATTTGCACGATGCGAAGACCTGTGAATGACCCACTTATATGAGATG 544
      |||::: |||::: |||::: |||::: |||::: |||:::
158  nLysPheValIThrGluValAlaProIleLeaProThrIleAspAlaGlyGly 175
      |||::: |||::: |||::: |||::: |||::: |||:::
545  GG...CGGAATATATCGATCAAAATTAATACCTCGACCGCTTGCGTAT 591
      |||::: |||::: |||::: |||::: |||::: |||:::
175  LysLeuAspThrTyrLysAspLysAsnArgPheSerSerPheValArgTyle 191
      |||::: |||::: |||::: |||::: |||::: |||:::
592  GGGGCGAGCGAGCA.....TATTGGCATGTGATGAAGATGACGCCCAA 635
      |||::: |||::: |||::: |||::: |||::: |||:::
192  GlyAlaGlyArgGlnLeuValTyrGlyLysGlyValTyrHisGlnGluG 208
      |||::: |||::: |||::: |||::: |||::: |||:::
636  TAACGCGCAAAATCATAT.....CATTTGCAAGTGGCGTTTCT 676
      |||::: |||::: |||::: |||::: |||::: |||:::
208  yAsn.....GluLysGlyTyrAspLeuAlaArgAspLeuSerGlnAlaTyrArgT 224

```

```

677 GCTGCTGGTGGCGG..... 690
224 yralaIleIaGlyThrProTylLysAspIleAsnIleAspGlnThrMet 240
691 AAATCCTTTGSCAAAAATGGATCAGTGTGGCAGATCAGTCAACTAGTAG 740
241 AsnThrGlnGlyLeuIleGlyPheGlyIAsnIleAsnIleGlyLysSerI 257
741 TGAAAAATTAAACATAGC...CCATAGTGTTTTAAACACAGAGAGT 787
257 AGIAGIleuLysGlnIleLeuSerGlnAspAlaIleThrAsnTyrGly 274
788 CATTTGGCAGCAGTGGCTCCACATGTTTAACTATGATGGCCAAAGCA 837
274 aIIeugLysPserGlySerProLeuPheIleAspLysGlnLysAsn 290
838 AAGTGTTAATTATGGGGATTTCGAACAGCGGACCCCTATTTAGAA 887
291 GlnThrValPheLeuGlyThrTyrAspTyrTyrAlaGlyTyrGlyLys 307
888 AAGCAATGGCTTCACAGTGGTGGTAAAGATTTGGTCTATGATGAATGT 937
307 sSerTyrGlnGlnThrAsnIleTyrLysGlnPheIleAspLysIle 323
938 TTGCTGGAGATCCCATTCAGTATTTACGAACACAGTCATAAATGGAAA 987
324 .....LysGlnHis..... 326
988 TACTCTTTAAGCAGCATATATATGGCAGAGAAATATCAATGCCAACAA 1037
327 .....AspAsnIleGlyThrValLysGlyAsnIleGlnI 338
1038 TGAACACATTCCTGCTCAATATGATTTAAACACAGCAACCTGTTCATTGT 1087
338 sHisTyrLysThr.....ThrGlyThr..... 345
1088 TTATATGTTTCTTATATCCGAGACAGCAAGAACTGTTTATCATCTGCA 1137
346 ..AsnSerHisIleGlySerThrAlaValArgLeuAlaAsnAsnGlnGly 361
1138 GGTGTGTCAACAGTATATGACCACAGTCAATATGAGAAATATTTTC 1187
362 AspAla.....AsnAsnGlyGlnAsnValThr 370
1188 CTTTATTATGAGCAAGAAAGCGAATTGATCTTATCCAGCAATCAATC 1237
370 rPhe.....GlnAspAsnIleThrLeuValLeuAsnGlnAsnIleAsnG 385
1238 AAGGTCTGTGAGATATATTTTCCAGAGGATTTTACGCTCGCGCTGAA 1287
385 IodIleIleGlyLeuPhePheIleLysGlyAspTyrThrValLysGlyAla 401
1288 AATAACGAA..ACTTGGCAAGCGCGGCGCTTCATCATAGTAAGACAG 1334
402 AsnAsnAspIleThrThrPheuGlyIleGlyIleAspValAlaAspGly 418
1335 TACCGTTACTTGGAAAGTAAAGCGGTGGCAGACAGCGCTGTCACAAA 1384
418 sLysValValIleThrGlnValLysAsnProAsnGlyAspArgLeuAlaLys 435
1385 TGGCAAAAGCAGCTGCACGTTCAAGCCAAAGGGGAAACCAAGAGCTCG 1434
435 IeGlyLysGlyThrLeuGlnIleAsnGlyThrGlyValAsnGlnGlyGln 451
1435 ATCAGCGTGGGAGAGTAAAGTATTTTGGATGACAGCAGCAGCAATA 1484
452 LeuLysValGlyAspGlyThrValIleLeuAsnGlnLysAlaAspAlaAs 468
1485 AGGCAAAAAACAGCCTTTAGTGAATTCGCTTGTCAGGCGCAGGGGTA 1534
468 pLysLysValGlnAlaPheSerGlnValGlyIleValSerGlyLysArgLys 485

```

```

1535 CGGTGCAATGAATGCCGATATCACTCAACCCGCAAACTCTATTTC 1584
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
485 hrleuValleuAsnSerSerAsnGlnIleasnProAspAsnLeuTyrphe 501
1585 GCGTTGCGCGCGGACGCTTGATTAACGGGCAATTCGCTTCGTTCA 1634
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
502 GtYrheTgTglYlArGleuAspAlaAsnGlyAsnSpreuThrheG 518
1635 CCGATATCAAAATACCGATAGAGGGGATGATTCACACCAATCAAG 1684
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
518 uHisIleArgAsnValasprGluGlyAlaArgIleValasnHisThr 535
1685 ACAAGATACACCGCTTACCATTAACGCAATGAAGATATGCTACAA 1734
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
535 sprHisAlaSerThrIleThrLeuThrGly...LysSerLeuIleThrAsn 550
1735 GGCATATCAACACGCTTGATAGC..... 1758
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
551 ProAsnSerLeuSerValHisSerIleGlnAsnAspTyrAspGluAsp 567
1759 ..... 1768
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
567 pTyrSerTyrTyrTyrArgProArgArgProIleProGlnGlyLysAsp 584
1769 TTGGCTAC..... 1776
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
584 euTyrTyrLysAsnTyrArgTyrTyrAlaLeuLysSerGlyGlyArgLeu 600
1776 ..... 1776
601 AsnAlaProMetProGluAsnGlyValAlaGluAsnAspTyrIlePh 617
1776 ..... 1776
617 emetGlyTyrThrGlnGlnGluAlaArgLysAsnAlaMetAsnHisLys 634
1777 ..... 1803
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
634 snAsnArgArgIleGlyAspRheGlyGlyRheRheAspGlnGluAsnGly 650
1804 ACCAAACGAAAGGGGCGCTACCTGTTTACACCGCCCGCGAGAGA 1853
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
651 LysGlnHisAsnGlyAlaLeuAsnLeuAsnRheAsnGlyLysSerAlaG 667
1854 CCGGACCGCTGCTTCGCGGCGGCAAAATTTAAACGCAACATCAAG 1903
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
667 nLysArgRheLeuLeuThrGlyAlaAsnLeuAsnGlyLysLieserV 684
1904 AAACAAAGCGCAAACTGTTTTCACCGGCGAGACCAACACCGCGCTAC 1953
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
684 alThrGlnGlyAsnValLeuLeuSerGlyArgProThrProHisAlaArg 700
1954 AATCATTTAAAGACCATTCGTCGCAAAAGAGGCG.....ATTCCTCG 1997
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
701 AspRheValAsnLysSerSerAlaArgLysAspAlaHisSpreSerLysAs 717
1998 CCGGGAATCTGTGGACAAACGATGATCAACCGCACATTTAAAGCG 2047
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
717 nAsnGluValAlaRheGluAspAspTyrIleAsnArgThrRheLysAla 734
2048 AAACCTTCCAAATTAAGCGGAGACGCGGCTTC.....CGCATGTT 2094
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
734 lAgIuIleValAlaAsnGlnSerAlaSerPheSerSerGlyArgAsnVal 750
2095 GCCAAAGTAAAGCGGATTCGATTCAGCATCAACCGCAAGCGATTT 2144
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
751 SerAspIleThrAlaAsnIleThrAlaTyrAspAsnAlaLysValAsnIle 767
2145 TGATGTCGACCGCATCAACGACACATCTGTACAGTTGCGAGTGA 2194
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
767 uGlyTyr.....LysAsnGlyAspGluValLysValArgSerAspTyrT 782
2195 CGGGTCTGACAAATGTGTGCAAAAAACCATTAACGATTAAGTATT 2244

```

```

      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
782 hrGlyTyrValThrCysAsnThrGlyAsnLeuSer...AspLysAlaLeu 797
2245 GCTTCATTGACTAAAGCGGACATCAAGCGCAATGCTGATCTGCGGATCA 2294
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
798 AsnSerRheAspAlaThrArgIleAsnGlyAsnValAsnLeu..... 811
2295 CCGTCATTTAAATGTCACAGGGCTTGCCACATCAACGGCAATCTAGTG 2344
811 ..... 811
2345 CAATGGGCAATACGTTATACATGTCACCCACAAACGCCCAAAAGCG 2394
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
812 ..... 813
2395 AACCTTACGCTGCGGCAATGCCAACACATTTAATCAAGCCATTT 2444
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
814 AsnAlaAlaLeuVal.....LeuGlyLysAlaAlaLe 824
2445 AAACGCAACACATCGCTTCGCGCAATGCTTCATTTAATCAAGCAC 2494
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
824 uThrGlyLysIleGlnGlnGlyAsnSerArgValSerLeuAsnGlnH 841
2495 ACGCGTCAAAACGCGACGTGACGCTTCGCGACAGCTTAAGCAAC 2544
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
841 Ls..... 841
2545 GTAAGCATTCGCACTCAACGGTAATGTCCTAGCGATTAAGCACT 2594
841 ..... 841
2595 ATTGCATTTTGAAGACGCGCTTACGCGCAATCAACGCGGCGCAAG 2644
841 ..... 841
2645 ATACGCAATTACATTAAGAACAGCAATGAGCGCTGCGCTACGACAG 2694
842 ..... 850
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2695 GAATGAGCAATTTAAACCTTGACAAAGCCACCATTTACATCAATTCG 2744
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
851 GlnValHisAsnLeuSerLeuAlaAspSerHisIleHisLeuAsnAl 867
2745 CTATGCCACGATGCGGCGCAAGCGCGCAACCGCACTGCGACAGATGCC 2794
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
867 a.....SerAspAlaGlnSerIle..... 873
2795 CCGGCGCGCGCTGCGCGCGCTGCGCGCTTCCTATTATCCGTTACAC 2844
873 ..... 873
2845 CCAACTTCGGTAGAATCCGCTTCAACAGCTGACGTTAAACGCGCAATT 2894
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
874 ..... 884
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2895 GAACGTCAGGAAACATTCGCTTATGTCGGAACCTTCGCTACGCGCA 2944
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
884 uSerGlyAsnGlyHisPheHisTyrLeuThrAspRheAlaLysAsnLeu 901
2945 GCGCAATGTAAGCTGCGGAAAGTTCGGAAGCACTTACACCTTCGCG 2994
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
901 LysAspLysValLeuValLysGlnSerAlaSerGlnHisTyrGlnHis 917
2995 GTCAACAATACGCGCAACGAACTGCAAGCTTCGGAACAATGACGGTAG 3044
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
918 ValGlnAsnLysThrGlyGlnPro...AsnGlnGlnGlyLeuAspRhe 933
3045 GGAAGGAAGACACAAACGCGCTGCGGAACCTTAATTTACACCTG 3094
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
933 eAspAlaSerSerValGlnAspArgSer...ArgLeuRheValSerLeu 949
3095 AAACGAAACAGTTCGATGCGCGCGCTGCGCTTACCAACTCATTCGCA 3144
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

949 laAsnHisTyrValAspLeuGluValLeuArgTyrThrIleLeuThrGlu 965
3145 GAGCGCGAGTTCGCGCTGCATAT.....CGGCT 3173
    ||||| ||||| |||||
966 AsnGlyIleThrAlaGlyLeuTyrAsnProTyrAlaGlyAsnGlyArgProVa 982
3174 CAAGAACACAGAGCTTTCGCAAACTCGCAGAGCAGAACCCAAAAAC 3223
    ||||| ||||| |||||
982 llyserProAlaProSerProAlaAlaAsnThrAlaSerGlnAlaGlnLysa 999
3224 AGCGGAAAAAGACACAGCCGCAAAAGCCTTCAGCGGCTGATTCGCGCGG 3273
    ||||| ||||| |||||
999 laThrGlnThrAspGlyAlaGln.....IleAlaLysPro 1010
3274 CGCGATGCGCTCGAAAAAGACAGAAAGCGTTCGACGCGCGC..... 3315
    ||||| ||||| |||||
1011 GlnAsnIleVal.....ValAlaProProSerProGlnAl 1022
3316 .....CGGACAGCAGCGCGGAAAAATGTCG 3340
    ||||| ||||| |||||
1022 aAsnGlnAlaGlnGluAlaLeuAlaArgGlnAlaLysAlaGlnValL 1039
3341 GCATTATCCAGCG.....GAGGA 3360
    ||||| ||||| |||||
1039 ysArgGlnGlnAlaAlaGlnAlaGlnLysValAlaArgGlnLysAspGln 1055
3361 GAGAAAAACGGGTGCAGCGCGAT..... 3384
    ||||| ||||| |||||
1056 GlnAlaLysArgLysAlaAlaGlnIleAlaArgGlnGlnGlnGlnAlaAr 1072
3385 .....AAG 3388
1072 gLysAlaAlaGlnLeuAlaAlaLysGlnLysAlaGlnAlaGlnArgLysa 1089
3389 ACACCGCTTCGCAACAGCGCGACAGCGAAACCGCGCGCTACACAC 3428
    ||||| ||||| |||||
1089 laArgGlnLeuAlaArgGlnLysAlaGlnLysAlaSerHisGlnAlaAsn 1105
3439 GCCTTCCCGCGCGCGCGCGCGCGCGGATTCGCGCA..... 3480
    ||||| ||||| |||||
1106 AlaLysProLysArgArgArgArgAlaIleLeuProArgProProAl 1122
3480 ..... 3480
1122 aProValPheSerLeuAspAspTyrAspAlaLysAspAsnSerGlnLeu 1139
3480 ..... 3480
1139 erIleGlyAsnLeuAlaArgValIleProArgMetGlyArgGlnLeuIle 1155
3480 ..... 3480
1156 AsnAspTyrGlnGlnIleProLeuGlnGlnLeuGlnAspGlnAlaGln 1172
3480 ..... 3480
1172 uGlnArgArgGlnAlaThrGlnPheHisSerLysSerArgAsnArgArg 1189
3480 ..... 3480
1189 laIleSerSerGlnProSerSerAspGlnAspAlaSerGlnSerValSer 1205
3480 ..... 3480
1206 ThrSerAspLysHisProGlnAspAsnThrGlnLeuHisGlnLysValGln 1222
3481 .....CTGCAACCCCAACCGGACCGCCCAAC 3505
    ||||| ||||| |||||
1222 uThrAlaGlnLeuGlnProArgAlaAlaGlnProArgThrGlnAlaAla 1239
3506 CGCAGCGCGACCTGATCAGCGTTATGCAATAGCGTTTACGGAATTT 3555
    ||||| ||||| |||||
1239 laGlnAlaAspAlaValSerThrAsnThrAsnSerAlaLeuSerAspAla 1255
3556 TCCGCAAGCTCAACAGCT.....TTCGCGTACA 3587
    ||||| ||||| |||||
1256 MetAlaSerThrGlnSerIleLeuLeuAspThrGlyAlaLysLeuThrAr 1272
3588 GAGCAATTAAGACCGCTATTTCGCAAGACCGCGCAACGCGCTTGA 3637
    ||||| ||||| |||||
1272 gHisIleAlaGlnLysSerArgAlaAspAlaGlnLysAsnSerValTPrM 1289
3638 CAACCGCATCCGGACACCAACACTACCTTCGCAAGATTTCCGCGC 3687
    ||||| ||||| |||||
1289 eTserAsnThrGlyTyrGlyArgAspTyrAlaSerAlaGlnTyrArgArg 1305
3688 TACCGCAACCAACCGACCTGCGCAATCGGATATGAGAAAAACCTCG 3737
    ||||| ||||| |||||
1306 PheSerSerLysArgThrGlnThrGlnIleGlyIleAspArgSerLeuSe 1322
3738 CAGCGGG...CGGTCGCGATCCTGTTTCGCAACCGGACCGAAACA 3784
    ||||| ||||| |||||
1322 rGlnAsnMetGlnIleGlyValLeuThrTyrSerAspSerGlnHisT 1339
3785 CTTTCACACAGCGCATCGCACTCGGACGCGCTTCGCGCGCGCGCTT 3834
    ||||| ||||| |||||
1339 hrPheAspGlnAla...GlyGlyLysAsnThrPheValGlnAlaAsnLeu 1354
3835 TTCGCGCAATACGCGATC...GACAGTTCTACATCGGATCGACGCGG 3881
    ||||| ||||| |||||
1355 TyrGlyLysTyrTyrLeuAsnAspAlaTyrValAlaGlyAspIleG 1371
3882 CGCGGCTTTTACGAGCGGACGCTTCAGACGCGATCGGCAAAATCC 3931
    ||||| ||||| |||||
1371 yAlaGlySerLeuArgSerArgLeuGlnThrGlnGlnLysAlaAsnPhe 1388
3932 GCGCGCGCGTGCATTCAGGATTCAGGACGACGATACCGCGCGTTTC 3981
    ||||| ||||| |||||
1388 snArgThrSerIleGlnThrGlyLeuThrLeuGlyAsnThrLeuLysIle 1404
3982 GCGGATTCGCGATTCGACCGCACATCGGCGCAACGCGTATTTCGTCA 4031
    ||||| ||||| |||||
1405 AsnGlnPheGlnIleValProSerAlaGlyIleArgTyrSerArgLeuSe 1421
4032 AAAGCGATTACCGCTACGAAAAAGCTCAATATCGCACCGCGCGCTTG 4081
    ||||| ||||| |||||
1421 rSerAlaAspTyrLysLeuGlyLysAspSerValLysSerMetAla 1438
4082 CATTCACCGCTACCGCGCGGCGATTAAGCAGATATCATTCACAACCG 4131
    ||||| ||||| |||||
1438 laValIlysthIleuThrAlaGlyLeuAspPheAlaTyrArgPheLys... 1453
4132 GCGCAACATTCATTCACACGCTTATTGACGCTGCTCATACGATGC 4181
    ||||| ||||| |||||
1454 ValGlyAsnLeuThrValLysProLeuLeuSerAlaAlaTyrPhe...Al 1469
4182 CGCTTCGCGCAAGCTCGCAACAGCGCTCAATACCGCGCTATTGCTCAG 4231
    ||||| ||||| |||||
1469 aAsnTyrGlyLysGlyGlyValAsnValGlyLysSerPheAlaTyrL 1486
4232 ATTTCGCAAAACCGCAGCGGAGTGGGCGTAAACCGCAATCAAA 4281
    ||||| ||||| |||||
1486 yAlaAspAsnGlnGlnGlnTyrSerAlaGlyValAlaLeuLeuTyrArg 1502
4282 GGTTCACGCTTCCTTCACAGCTGCGCGCGCGCAAGCGCGCATGGA 4331
    ||||| ||||| |||||
1503 AsnValThrLeuAsnValAsnGlySerIleThrLysGlyLysGlnLeuG 1519
4332 AGCGCAACACAGCGCGCGCATCAATTAAGCATACGCGCTG 4371
    ||||| ||||| |||||
1519 uLysGlnLysSerGlyGlnIleLysIleGlnIleArgPhe 1532

```

seq_name: SwissProt_40:IGA0_HAEN

seq_documentation_block:

ID IGA0_HAEN STANDARD:

PRT; 1694 AA.


```
718 GTGGCAGCTCACTAGTAGTGAATAAATTAACATAGC...CCATA 764
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
255 ASnGlyLeuIleGlyPheGlyAsnSerLysGluGlnHisSerAspProly 271
765 TGGTTTTTTTACCAACAGGA.....GCCTATTGGCGACA 799
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrlaValLeuGlyAspS 288
800 GTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGTAAT 849
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
288 erGlySerProLeuPheValTyrlAspArgGluLysGlyLysIlePhe 304
850 AATGGGTATTGCAACGGGCAACCCCTATATAGAAAAGCAATGGCTT 899
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
305 LeuGlySerTyrlAspPheTyrPalaglyTyrlAsnLysLysSerTyr 321
900 CCAGCTGGTTCGTAAAGATTGGTCTATGATGAATCTTGGCGAGATA 949
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
321 urTyrAsnIleTyrlLysProGluPheAlaLysThrValLeuAspLys 338
950 CCCATTCACTATTCTAGCAACACGTCAAATGGAAATACCTTTTAA 999
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
338 hrAlaGlySerLeu.....ThrGlySerAsnThrGlnTyrlAsnTyr 352
1000 GACGATAATTAATGGCACAGAAAATCAATGCCAAACATGAACACA 1049
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
353 Pro.....ThrGlyLysThrSerVal..... 359
1050 TCTGCTAATAGATTAAACACGACCGCTTCATTTGTTTATGTTCTT 1099
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
360 .IleSerAsnGlySerGluSerLeuAsnValAspLeuPheAspSerSe 376
1100 TATCCAGACAGCAGACAGAACTGTTTATCATGCTGCAGGTGTGTC 1149
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
376 IAspThrAspSerLys..... 381
1150 AGTTATCCAGCCAGACTGATAATAGAGAAATATTCTTTATTTAGCA 1199
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
382 .....LysAsnAsnHisGlyLysSerValThrLeu.....Ar 392
1200 AGGAAAGCGCAATTGATCTTACCGACACATCATCAAGTGGCGGAG 1249
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
392 gGlySerGlyThrLeuThrLeuAsnAsnHisGlnGlyLysGly 409
1250 GATTATATTTCCAAAGAGATTTTACGGTC...TCGGCTGAAAATAAC 1296
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
409 LysLeuPhePheGluGlyAspTyrGluValLysGlyThrSerAspSer 425
1297 ACTTGGCAGGGCGGCGCTTCATATCACTAGAGACAGTACCGTTACT 1346
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
426 ThrTyrLysGlyAlaGlyValSerValAlaAspGlyLysThrValTh 442
1347 GAAAGTAAACGGCGTGGCAACGACCGCTGTCCAAAATCGGCAAGGA 1396
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
442 PylsValHisAsnProLysSerAspArgLeuAlaLysIleGlyLysG 459
1397 CGCTGACGCTTAAGCCAAAGGGGAAAACCAAGCTCGATCAGCGTGG 1446
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
459 hrLeuIleValGluGlyLysGlyLysAsnLysGlySerLeuLysVal 475
1447 GACGGTACGTCATTTTGGATGACGAGCAGACGATTAAGGCAAAAA 1496
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
476 AspGlyThrValIleLeuLysGlnGlnAlaAspAlaAsnAsnLys 492
1497 AGCCTTTAGTGAATCGCTTGGTCAGCGGAGGAGTACGGTCAACTGA 1546
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
492 sAlaPheSerGlnValGlyIleValSerGlyArgSerThrValValLe 509
1547 ATGCCGATATATGATTCAACCCCGCAAACTCTATTTGGCTTGGCGGC 1596
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
509 snAspAspLysGlnValAspProAsnSerIleTyrlPheGlyPheArg 525
1597 GACGCTTTGATTTAAACGGGCAATTCGCTTTCGTCACCGATTTCAA 1646
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
526 GylAlaGlyLeuAspAlaAsnGlyAsnAsnLeuThrPheGlnHisIle 542
1647 TACCGATGAAGGGCGATGATTGTCAACCAATCAAGACCAAGATCCA 1696
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
542 nIleAspAspGlyAlaIleArgLeuValAsnHisAsnThrSerLysTh 559
1697 CCGTTACCATTTACAGC..... 1713
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
559 hrValThrIleThrGlyLysLeuIleThrAspProAsnThrIleThr 575
1713 ..... 1713
576 ProTyrlAsnIleAspAlaPheAspGluAspAsnProTyrlAlaPhe 592
1713 ..... 1713
592 GileLysAspGlyGlyGlnLeuTyrlLeuAsnLeuGluAsnTyrlThr 609
1713 ..... 1713
609 yrlAlaLeuArgLysGlyAlaSerThrArgSerGluLeuProLysAsn 625
1714 .....ATAAAGATATTGC 1727
626 GlyLysSerAsnGluAsnTyrLeuTyrlMetGlyLysThrSerAsp 642
1728 TACAACCGGCAATTAACAACAGCTTGATGCAAAAAGAAATTCCTACA 1777
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
642 aLysArgAsnValIleMetAsnHisIleAsnAsnGluArgMetAsnG 659
1778 ACGGTGGTGTGGGAGAAAGATACAGCAAAAGCAAGCGGGCGCTAAC 1827
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
659 snGlyTyrlPheIleGlyLeuGlu...GlyLysAsnAsnGlyAsnLe 674
1828 CTGTGTTTACAGCCCGCGGAGAGACCGCACCTGCTGCTTCCGGCGG 1877
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
675 ValThrPheLysGlyLysSerGluGlnAsnArgPheLeuLeuThrGl 691
1878 AACAAATTTTAAACGGCAACATCACGCAAAACCAACGCAAACTGTT 1927
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
691 yThrAsnLeuAsnGlyAspLeuLysValGlyLysGlyThrLeuPhe 708
1928 GCGCAGACCAACACCGACCGCTTACAT..... 1956
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
708 erGlyArgProThrProHisAlaArgAspIleAlaGlyIleSerSer 724
1957 CATTTAAACGACATTTGGTCGCAAAAAGAGGCAATTCCTCGCGGGA 2006
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
725 LysLysAspArgLysPheAlaGlu.....AsnAsnGluVal 736
2007 CGTGTGGACACGACATGTGATCAACCGCACATTTAAAGCGGAAATCC 2056
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
736 IValIleValGluAspAspTyrIleAsnArgAsnPheLysAlaThrAs 753
2057 AAATTTAAAGGCGAGCGAGCGGTGCTTC...CGCAATGTTGCCAAAG 2103
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
753 snValThrAsnAsnAlaThrLeuTyrlSerGlyArgAsnValAlaAs 769
2104 AAAGCGCATTTGGCATTTAGCAATCACGCCCAAGCACTTTTGGTGC 2153
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
770 ThrSerAsnIleThrAlaSerAspAsnAlaLysValHisIleGly 785
2154 ACCGCATCAAAAGCCACAACTGTATACGTTGGATGGAGCGGCTCA 2203
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
786 ...LysAlaGlyAspThrValLysValArgSerAspTyrThrGly 801
2204 CAATTTGTGCGAAAAAACCATTTACCGAGATTAAGGATGCTTCATTG 2253
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
801 alThrGlySerThrThrAspLysLeuSer...AspLysAlaLeuAsn 816
2254 ACTAAGCCGACATACGGGCAATGTGATCTTCCGATCACCGTCAATT 2303
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

817 AsnAlaThrAsnValSerGlyAsnValAsnLeuSer 828
2304 AAATCTCAGAGGGCTTGCACACTCAGCGCAATCTTAGTCAAAATGGCG 2353
828 828
2354 ATACAGCTTATACAGTACGCCAACAGCCCAAAAGCGCACTTACG 2403
828 828
2404 CTCTGGGCAATGCCCAAGCAACATTATATCAGCCACATTAAAGCGCA 2453
829GlyAsnAlaAsnPhenValLeuGlyLysAlaAsnLeuPheGlyTh 843
2454 CACATCGCCTTGGGCAATGCTTCAATTATCTAAGCAGCAGCCGCTAC 2503
843 rIleSerGlyThrGlyAsnSerGlnValAlaArgLeuThrGlu..... 856
2504 AAAAGCGCAGCTGACGGCTTCCGGCAACGCTAAGCAAGTAAGCCAT 2553
856 856
2554 TCCGCACTCAACGTAATGTCTCCTAGCGATAAGCAGTATTCATTT 2603
856 856
2604 TGAAGACAGCCGCTTACCGGCAATCAGCGCGGCAAGATACGGCAT 2653
856 856
2654 TACACTTAAAGACAGCGAATGACGCTGCCCTCAGCAGCAGTAATGAGC 2703
857AsnSerHisThrPheIleuThrGlyAspSerAsnValAsn 869
2704 AATTTAACCTTGGACACGCCACCTTACACTCAATTCGCTTACGCCCA 2753
870 GlnLeuAsnLeuAspLysGlyHisIleHisLeuAsnAla.....GlnAs 884
2754 CGATGGGCGAGGGGCGCAACCGCGCAGTGCAGACAGATGGCGCGCGCCG 2803
884 nAspAlaAsnLysValThrThr..... 891
2804 GTTCGGCGCTTCGGCGCGCTCCATTATTCGTTACACCGCAACTTCG 2853
891 891
2854 GTAGAATCCCGTTTCAACAGCGTGAACGTAACGGCAATTAAGGTCAC 2903
892TyrAsnThrIleuThrValAsn...SerLeuSerGlyAs 903
2904 GGAACATTCGCTTATGTGCGAACTCTTCGCTACCGCAGCAGCAAT 2953
903 nGlySerPheTyrTyrIleuThrAspLeuSerAsnLysGlnGlyAspLysV 920
2954 TGAAGCTGCGGAAAGTTCGGAAGCATTACACTTGGCGGTCAACAT 3003
920 aIValValThrLysSerAlaThrGlyAsnPhenThrLeuGlnValAlaAsp 936
3004 ACCGGCAAGACCTGCAAGCCTCGAAACAATTGACGCTAGTGAAGGAAA 3053
937 LysThrGlyGlnProThrLys...AsnGlnLeuThrLeuPheAspAla... 951
3054 AGACAACAACCGCTGTCCGAAAACTTAATTTCACCTCGCAAAAAGAAC 3103
952SerAsnAlaThrAlaGlnAsnLeuAsnValSerLeuValGlyAsnT 967
3104 ACCTGCATGCGCGCGCTGACCACTATCCGCAAGAGCGCGGAG 3153
967 hrValAspLeuGlyAlaThrLysTyrLysLeuArgAsnValAsnGlyArg 983
3154 TTCCGCTGCATTAATCG..... 3171
984 TyrAspLeuTyrAsnProGlnValGlyLysArgAsnGlnThrValAspTh 1000

3171 3171
1000 rThrAsnIleThrThrProAsnAsnIleGlnAlaAspValProSerValP 1017
3171 3171
1017 roSerAsnAsnGlnGluIleAlaArgValGluThrProValProPro 1033
3171 3171
1034 AlaProAlaThrProSerGluThrThrGluThrValAlaGluAsnSerLy 1050
3171 3171
1050 sGlnGluSerLysThrValGluLysAsnGlnAspAlaThrGluThrT 1067
3171 3171
1067 hrAlaGlnAsnGlyGluValAlaGluGluAlaLysProSerValLysAla 1083
3171 3171
1084 AsnThrGlnThrAsnGlnValAlaGlnSerGlySerGluThrGluGluTh 1100
3172GTCAAAGAA..... 3180
1100 rGlnThrThrGluIleLysGluThrAlaLysValGluLysGluLysAla 1117
3181CAAGAG 3186
1117 IalysValGluLysAspGluIleGlnGluAlaProGlnMetAlaSerGlu 1133
3187 CTTTCCGACAACTCGGCAAG..... 3207
1134 ThrSerProLysGlnAlaLysProAlaProLysGluValSerThrAspTh 1150
3207 3207
1150 rLysValGluGluThrGlnValGlnAlaGlnProGlnThrGlnSerThrT 1167
3208GCAGAGCC 3216
1167 hrValAlaAlaGluAlaThrSerProAsnSerLysProAlaGluGlu 1183
3217 AAAAAGACGCGGAAAAAGACAGCGCGCAAAAGCTTGACGCGTGTGC 3266
1184 ThrGlnProSerGluLysThrAsnAlaGluProValThrProValAlaSe 1200
3267 G.....G 3268
1200 rLysAsnGlnThrGluAsnThrThrAspGlnProThrGluArgGluLysT 1217
3269 CCGGGCGCGATCGCGTGAAGAGAGAGAGCGTTCGCAAGCGCGCGG 3318
1217 hrAlaLysValGluThrGluLysThrGlnGluProProGlnValAlaSer 1233
3319 CAGCGCAGGC..... 3327
1234 GlnAlaSerProLysGlnGluGlnSerGluThrValGlnProGlnAlaVala 1250
3328GGGGAATATGTC..GGCATTATGACGCGGAGAGAGAGAAA 3367
1250 IleuGlnSerGluAsnValProThrValAsnAsnAlaGluGluValGlnAla 1267
3368 AAGGGGTGCAGCGCATTAAGAC..... 3390
1267 IagInLeuGlnThrGlnThrSerAlaThrValSerThrLysGlnProAla 1283
3391ACCGCTTGGCGAAACAGCG 3410
1284 ProGluAsnSerIleAsnThrGlySerAlaThrAlaThrGluThrAla 1300

[illegible]

```

seq_name: SwissProt_40:IGA2_HAEIN
seq_documentation_block:
ID   IGA2_HAEIN      STANDARD;          PRT;   1702 AA.
AC   P45384;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN   IGA.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxId=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HK715 / SEROTYPE B;
RA   MEDLINE=92234949; PubMed=1373717;
RX   Foulson K., Reinholdt J., Kilian M.;
RT   "A comparative genetic study of serologically distinct Haemophilus
RT   influenzae type 1 immunoglobulin A1 proteases.";
RL   J. Bacteriol. 174:2913-2921(1992).
CC   - FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC   PRODUCING INTACT FC AND FAB FRAGMENTS.
CC   - CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
CC   certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC   substrates are known.
CC   - SUBCELLULAR LOCATION: Secreted.
CC   - DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC   SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC   OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC   DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC   SIMILARITY).
CC   - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASES).
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb.ch/announce/
CC   or send an email to license@isb.slb.ch).
-----
DR   EMBL, M87489; AAA24966.1; -.
DR   MEROPS, S06.001; -.
DR   InterPro: IPR000710; IGA_S6.
DR   Pfam: PF02395; IGA1; 1
DR   PRINTS, PR00921; IGASEPR7ASE.
KW   Hydrolase; Serine protease; Transmembrane; Zymogen; Repeat; Signal.
FT   SIGNAL          1      25
FT   POTENTIAL.

```


| | | | |
|------|---|------|-------|
| 2204 | CAAAATGTGTGCAAAAAAACCATTCACGAGCATTAAGATGATTCCTCATGT | 2255 | |
| 801 | alThrCysThrThrAspLysLeuSer... AspLysAlaLeuSerPhe | 816 | |
| 2254 | ACTTAGACCGACATCAGCGGCAATGTGCATCTTCCGATCAGCGTCATTT | 2303 | |
| 817 | AsnAlaThrAsnValSerGlyAsnValAsnLeuSer..... | 828 | |
| 2304 | AAATCTCAGAGGGCTTGCCACACTCAGCGCATCTTAAGTCAAAATGGCG | 2353 | |
| 828 | | 828 | |
| 2354 | ATACAGCTTATACAGTCAGCCACAAAGCCAAACCAAAAGCGAACTTAGC | 2403 | |
| 828 | | 828 | |
| 2404 | CTGTGGGCAATGCCCAAGCACAATTAAATCAAGCCACATTAAGCGCAA | 2453 | |
| 829 | GlyAsnAlaAsnPheValLeuGlySalAlaAsnLeuPheGlyTh | 843 | |
| 2454 | CACATCGGCTGGGGAATGCTTCATTATTAAGTAAAGCGACCGCGGTAC | 2503 | |
| 843 | rlIleSerGlyThrGlyAsnSerGlnValArgLeuThrGlu..... | 856 | |
| 2504 | AAAAGGCAAGTCTGACGCTTCCGGCAACGCTAAGCAAAAGTAAGCCAT | 2553 | |
| 856 | | 856 | |
| 2554 | TCCGCACTCAAGCGTAATGTCTTCCTAGCCGATAAGCAGTAATTCATTT | 2603 | |
| 856 | | 856 | |
| 2604 | TGAAGCGACGCGCTTACCGACAATCAGCGCGGCAAGATACGGCAT | 2653 | |
| 856 | | 856 | |
| 2654 | TACACTTAABAACAGCGAATGGACGCTGCCGTACGACGCGAATTAGGC | 2703 | |
| 857 | AsnSerHisIleuThrGlnGlyAspSerAsnValAsn | 869 | |
| 2704 | AAATTAACCTTGAACAGCGCACCATTCACACCAATTCGGCTTAAGCGCA | 2753 | |
| 870 | GlnLeuAsnLeuAspLysGlyHisIleHisLeuAsnAla..... GlnAs | 884 | |
| 2754 | CGATCGCGCAGGCGCGCAACCGGCGATGGCAGATGCGCGCGCGCC | 2803 | |
| 884 | nasPAlaAsnLysValIleThr..... | 891 | |
| 2804 | GTTCCGCGCGTTCCGGCGTTCCATTAATTCGTTACACCGCCAACTTCG | 2853 | |
| 891 | | 891 | |
| 2854 | GTAGAAATCCCGTTTCAACAGCGGTGAACGGCAATTTGAACGGGTCA | 2903 | |
| 892 | TyrAsnThrLeuThrValAsn... SerLeuSerGlyAs | 903 | |
| 2904 | GGGACATTCGCGTTTAATGTGGGAACCTTCGCTACCGCAGCAGCAAAAT | 2953 | |
| 903 | ngLysSerPheIleuThrLeuThrAspLeuSerAsnLysGlnGlyAspLysV | 920 | |
| 2954 | TGAACTGGCGGAAAGTTCCGAGAGCACTTACACCTTGGGCGTCAACAAT | 3003 | |
| 920 | alValIleThrLysSerIleThrAlaThrGlyAsnPheThrLeuGlnValAlaAsp | 936 | |
| 3004 | ACCGGCAACGAACCTGCGAAGCTCGACCAATTTAGCGGTAGTGGAGAGAA | 3053 | |
| 937 | LysThrIleGlyLupProIleLys... AsnGlnLeuThrLeuPheAspAla... 951 | | |
| 3054 | AGACACAACAACCGCTGTCCGAAAACCTTAATTCACCTTCGCAAAAGAAC | 3103 | |
| 952 | SerAsnIleThrArgAsnAsnLeuAsnValSerLeuValGlyAsnT | 967 | |

```
3104 ACCTGATGCGGCGGCTTACCACTACCGCAAGAGCGGAG 3153
      ||||| ||||| ||||| ||||| ||||| |||||
967 hrValAspLeuGluAlaIrrPylsLeuArgAsnValAsnGlyArg 963
3154 TTCGCGCTGCATATCCG..... 3171
      :: ||||| ||||| ||||| ||||| |||||
984 TyrAspLeuTyrAsnProGluValGluLysArgAsnGlnThrValAspTh 1000
3171 ..... 3171
1000 rThrAsnIleThrThrProAsnAsnIleGlnAlaAspValProSerValP 1017
3171 ..... 3171
1017 roSerAsnAsnGluGluIleAlaArgValGluThrProValProProPro 1033
3171 ..... 3171
1034 AlaProAlaThrProSerGluThrThrGluThrValAlaGluAsnSerLy 1050
3171 ..... 3171
1050 sGlnGluSerLyThrValGluLysAsnGluGlnAspAlaThrGluThrT 1067
3171 ..... 3171
1067 hrAlaGlnAsnGlyGluValAlaGluGluAlaLysProSerValLysAla 1083
3171 ..... 3171
1084 AsnThrGlnThrAsnGluValAlaGlnSerGlySerGluThrGluGluTh 1100
3172 .....GTCAAGAACAGAGCTTCCGACAACTCGCAGAG 3208
      ::||| ||||| ||||| ||||| |||||
1100 rGlnThrThrGluIleLysGluThrAlaLysValGluLysGluGluLysA 1117
3209 CAGAAAGCAAAAAACAG..... 3225
      ||||| ||||| ||||| ||||| |||||
1117 IalysValGluLysGluLysAlaLysValGluLysAspGluIleGln 1133
3225 ..... 3225
1134 GluAlaProGlnMetAlaSerGluThrSerProLysGlnAlaLysProAl 1150
3225 ..... 3225
1150 aProLysGluValSerThrAspThrLysValGluGluThrGlnValGlnA 1167
3225 ..... 3225
1167 IagInProGlnThrGlnSerThrThrValAlaAlaAlaGluAlaThrSer 1183
3226 .....CGCAAAAAAGCAACGC 3242
      ::||| ||||| ||||| ||||| |||||
1184 ProAsnSerLysProAlaGluGluThrGlnProSerGluLysThrAsnAl 1200
3243 GCAAAGCCTTGACGCGCTGATTGCG..... 3267
      ::||| ||||| ||||| ||||| |||||
1200 aGluProValThrProValIleSerLysAsnGlnThrGluAsnThrThrA 1217
3268 .....GCGGGCGCGATGCGTGAAGAGCA 3294
      ::||| ||||| ||||| ||||| |||||
1217 sPglInProThrGluArgGluLysThrAlaLysValGluThrGluLysThr 1233
3295 GAAAGCGTTGCGAGACCGCCGCGAGCAGC..... 3327
      ::||| ||||| ||||| ||||| |||||
1234 GlnGluProProGlnValAlaSerGlnAlaSerProLysGlnGluInSe 1250
3328 .....GGGAAATGTC...GGCA 3343
      ::||| ||||| ||||| ||||| |||||
1250 rGluThrValGlnProGlnAlaValLeuGluSerGluAsnValProThrV 1267
3344 TTATCGAGCGAGAGAGAAAAAAGCGGTGCGAGCGGATTAAGAC... 3390

:: ::||| ||||| ||||| ||||| |||||
1267 aIAsnAlaGluGluValGlnAlaGlnLeuGlnThrGlnThrSerAla 1283
3390 ..... 3390
1284 ThrValSerThrLysGlnProAlaProGluAsnSerIleAsnThrGlySe 1300
3391 ...ACCGCTTGCGGAAACAGCGGAGCGGAAACCGCGCGGTACC. 3435
      ||||| ||||| ||||| ||||| |||||
1300 rAlaThrAlaIleThrGluThrAlaGluLysSerAspLysProGlnThrG 1317
3435 ..... 3435
1317 IuThrAlaAlaSerThrGluAspAlaSerGlnHisLysAlaAsnThrVal 1333
3436 .....ACCGCTTCCCGCGCGCCG 3455
      ::||| ||||| ||||| ||||| |||||
1334 AlaAspAsnSerValAlaAsnAsnSerGluSerGluProLysSerAr 1350
3456 CCGCGCGCGCGGATTTGCGCACTGCA..... 3486
      ||||| ||||| ||||| ||||| |||||
1350 garG...ArgArgSerIleSerGlnProGlnGluThrSerAlaGluGluT 1366
3486 ..... 3486
1366 hrThrAlaAlaSerThrAspGluThrThrIleAlaAspAsnSerLysArg 1382
3487 CCCCACCGCAGCCCGCACCGCAG..... 3510
      ::||| ||||| ||||| ||||| |||||
1383 SerLysProAsnArgArgSerArgArgSerValArgSerGluProThrVa 1399
3511 .....CGCGACCTGATCAGCC 3526
      ||||| ||||| ||||| ||||| |||||
1399 IThrAsnGlySerAspArgSerThrValAlaLeuThrAspLeuThrSerT 1416
3527 GTTATGCCATAGCGTTTGAGTGAATTTCCGCC..... 3561
      ::||| ||||| ||||| ||||| |||||
1416 hrAsnThrAsnAlaValIleSerAspAlaMetAlaLysAlaGlnPheVal 1432
3562 ACGCTCAACAGCGTTTGCCTGACGACAGCAATTAGACCGGTA...TT 3608
      ::||| ||||| ||||| ||||| |||||
1433 AlaLeuAsnValGluLysAlaValSerGlnHisIleSerGlnLeuGluMe 1449
3609 TGCCGAAGACCGCGGACAGCGCTTGACAGCGGACATCCGGACACCA 3658
      ::||| ||||| ||||| ||||| |||||
1449 IAsnAsnGluGluGlnThrAsnValThrPalaSerAsnThrSerMetAsnG 1466
3659 AACACTACCGCTTGCAAGATTTCCGCGCTACCGCCCAACACCGACCTG 3708
      ::||| ||||| ||||| ||||| |||||
1466 IuAsnThrSerSerSerGlnThrArgArgPheSerSerLysThrGln 1482
3709 CGCCAAATCGGTATGCAGAAAAACTCGGACGCGG...CGCGTCGGCAT 3755
      ||||| ||||| ||||| ||||| |||||
1483 ThrGlnLeuGlnThrAspGlnThrIleSerAsnAsnValGlnLeuGly 1499
3756 CCGTTTTCGACACACCGGCGGAAACACTTCGACGACGACGATCGGA 3805
      ::||| ||||| ||||| ||||| |||||
1499 yValPheThrThrValArgAsnSerAsnAsnPheAspLysAlaSerSerL 1516
3806 ACTCGGACGCGTTGCCAGCGCGCGCTTTCGCGCAATACGGCATGAC 3855
      ::||| ||||| ||||| ||||| |||||
1516 yAsnThr...LeuAlaGlnValAlaAsnPheThrSerLysThrAlaAsp 1531
3856 ...AGTTTACATCGGATCAGCGCGCGCGGCTTTAGCAGCGGAG 3902
      ::||| ||||| ||||| ||||| |||||
1532 AsnHisThrPylLeuGlyIleAspLeuGlyGlyLysPheLeuInSerAs 1548
3903 CCGTTGACGCGATGAGGAGCAAAATCGCGCGCGCTGCTCATTAAG 3952
      ::||| ||||| ||||| ||||| |||||
1548 pLeuLysThrAsnHisAsnAlaLysPheAlaGlnHisThrAlaGlnPheG 1565
3953 GCATTGACGACGATACCGCGCGGCTTTCGCGGATTCGACATCGAACCG 4002
      ||||| ||||| ||||| ||||| |||||
```

```

1565 lyleuThralaglyLysalaPheasnleuGlyasnphgilylerthrPro 1581
4003 CACATCGCGCCGACGCGCTATTGTCACAAAGCGATTACCGCTTCGA 4052
1582 lIleValGlyValArgTyrSerTyrLeuSerAsnAlaSnPheAlaLeuAl 1598
4053 AAACGTCAATATCGCCACCCCGGCTTGATTCACCGCTACCGCGG 4102
1598 alyAspArgIleLysValAsnProIleSerVallySThrAlaPheAlaG 1615
4103 GCATTAGGCGAGTATTATTCATCAACCGCGCACACATTCATTCACG 4152
1615 lnyAlaPleuSerTyrThrHisLeuGlyGlu...PheSerValThr 1630
4153 CCTATTGTGAGCGCTGCTCAT...ACCGATGCGCGCTGCGGCAAGCGG 4199
1631 ProIleuSerAlaArgTyrAspThrAsnGlnGlySerGlyLysIle.. 1646
4200 AACACGCGTCAATACCGCGCTATTGCTCAGATTCGCGAAACCGCGA 4249
1647 ...AsnValAsnGlnTyrAspPheAlaTyrAsnValGluAsnGlnG 1662
4250 GTCCGATGCGGCGGTAAACGCGCAATCAAGTTTCACGCTGCTCC 4299
1662 lnyTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSerLeu 1678
4300 CACGCTGCGCGCGCAAGCGCGCACTGAGAGCGCAACACACGCGG 4349
1679 lIleGlyGlyLeuThrLysAlaLysGlnAlaGlyLysGlnLysThrAlaG 1695
4350 CATCAATATGAGCTAC 4365
1695 uleuLysLeuSerPhe 1700

seq_name: SwissProt_40:IGA4_HAEIN

seq_documentation_block:
ID IGA4_HAEIN STANDARD: PRT: 1849 AA.
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NHTI HK61;
RX MEDLINE-92234949; PubMed-1373717;
RA Poulson K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTERACT EC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE. AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M87491; AAA24968.1;
DR MEROPS; S06.001;
DR InterPro; IPR000710; IGA-S6.
DR Pfam; PF02395; IGA1; 1.
DR PRINTS; PR00921; IGASERPTASE.
KW Hydrolyase; Serine protease; Transmembrane; zymogen; signal.
FT SIGNAL 1
FT CHAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1022 1849 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 299 299 PROBABLE.
FT SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AFA CRC64;

alignment_scores:
Quality: 1488.50 Length: 1917
Ratio: 1.493 Gaps: 59
Percent Similarity: 52.008 Percent Identity: 24.935

alignment_block:
US-09-303-518D-649 x IGA4_HAEIN
Align seg 1/1 to: IGA4_HAEIN from: 1 to: 1849

64 CGCTTCTGCGCTTACTTACCATATGCTGCTGCTGCGCAATTCCTCC 113
5 LysPheLysLeuAsnPheAlaAlaLeuThrValAlaTyrAlaLeuThrPr 21
114 CCAAGCTTGGCGGCGACACTATTTCGCGATCACTACCAATACATATC 163
21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38
164 GCGACTTTCGCGAAATTAAGCAAGTTTCGAGTCGGGCGAAAGATATT 213
38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54
214 GAGGTTTCAACAAAAAGGAGAGTTGGTCGCGCAATCATGACAAAGC 263
55 GluValArgAspLysLysAsnGlnSerLeuLysGlnAlaLeuProAsnG 71
264 C...CCGATGATGATTTTCTGTGTGTG...CGTAAGCGCGTGGCGG 307
71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88
308 CATTGTGGCGGATCAATATATTGTGAGCGTGGACAT...AACGCGGC 354
88 hrLeuValAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
355 TATAACAACGTGATTTTGTG...GCGGAGAGAGAAA 389
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnGlnGly 121
390 TCCGATCAACATCGTTTACT...TATAAATG 421
121 nAlaLysSerHisArgAspValSerSerGluGluAsnArgTyrTyrThy 138
422 TGAACGCAATTAATT...AAA 441
138 alGluLysAsnAsnPheProThrGluAsnValThrSerPheThrLys 154
442 GCAAGGCAATAAGGCCATCTTATGGGCGGATTAATCATATCGCGGTT 491
155 GluGluGlnAspAlaGlnLysArgArgGluAspTyrThrProArgIle 171
492 GCATTAATTTGTCAAGATCAGACACCTGTTAAAGACAGTTATATG 541
171 uAspLysPheValThrGluValAlaProIleGluAlaSerThrAlaAsn 188
542 ATGGCGG...AATATATCATCAAAATTAATTAACCTGACCGTGTGCT 588

```

```
188 snbsnlysglylutyraasnaserasplyslyrproalphevalarg 204
589 ATTGGGCGAGGAGGCAATAT..... 609
205 leuglyserglyserglnpheiletyrlysglyserargtyrcline 221
610 .TGCGCATCTGATGAGATGAGCCCAATACCCGGAAGTTCATATCAT. 657
221 uileuethrclulysasplysglnleuaualeuargasntrpsrv 238
658 .....ATTGCAAGTCGCTATCTTGCGCTGTT 684
238 algllyasrpsnleuclleuvalgllyasnlaetyrthrlyrglylle 254
685 GGTGGCAATACCTTTGCAAAATGATGATGAGGTGGGACAGTCAACTT 734
255 Alaglythrprotyrlyrsvalasnhisgluasnbnllyleuileglyph 271
735 AGGTAGTGAATAATTAACATAGC...CCATATGCTTTTTCACCAACAG 781
271 eglyasnserlysglulhisserasprplyslylleuenserclna 288
782 GA.....GGTCATTTGGCGAGAGTGGCTCAGCAATGTTT 816
288 sprproleuthrasntyrAlaValleucllyaspseryserproleuphe 304
817 ATGTATGATGCCAAAGCAAAAGTGGTAAATTAATGCGGTATTTGCAAC 866
305 Valtyrasparglyllysglysrpleupheleucllysertryasrph 321
867 GGGCAACCCCTATATAGGAAAGCAATGGCTTCAGCTGGTGTGTAAG 916
321 etrplaglytryasnlylsysetrtpglnclutrpasnilletyrlysh 338
917 ATTGGTCTATGATGAATCTTTGCTGGAGATACCATTCAGTATCTAC 966
338 lsglnphealaglyllyser.....GlnGlnTyrSerAlaglyser 352
967 CAACCACTCAAAATGGAATACTCTTTTAACGACGATATATATGGCAC 1016
353 leuthrlyserasnthrGlnTyrThrTp.....GlnAlaIth 365
1017 AGGAAATCAATGCCAAACATGAACAATCTCTGCTAATGATTA 1066
365 rgly.....S 367
1067 AAACAGAACCGTT.....CAATTGTTAATGTTCTTTATCC 1104
367 etthrserthrlethrcllysglyluproleuSerValaspleuthr 383
1105 GAGACAGCAAGAGACCTGTTATCATGCTGAGGTGTCACACAGTTA 1154
384 aspglylyasrplysPro..... 389
1155 TCGACCCAGACTGAATAATGAGAAATATTTCTTTATGACGAGGAA 1204
390 .....AsnHisglySerlethrleu.....lysglys 400
1205 AAGCGAATGCTACTTACGACCAACATCAAGGTGGGAGAGATTA 1254
400 erglythrleuthrleuasnbnhisleaspglnlyalaglyllyleu 416
1255 TATTTCAAGAGATTTTACGGTC...TCGCTGAAATTAACGAACCTTG 1301
417 phepheglnuglyasprtyrclvallysglythrseraspsrthrThr 433
1302 GCAAGCGCGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGAAG 1351
433 plysglyalaglyvalserValaIasprlylysrThrValaThrTrplysv 450
1352 TTAACGGCTGCAAAAGCGCGCTGTCCAAATTCGCAAAAGCGAGCTG 1401
450 alHisasnProlystytraspargleuAlaValleucllysglythrleu 466

1402 CACGTTCAAGCCAAAGGGGAAACCAAGCTCGATCAGCGTGGCGAGG 1451
467 ValValGlnGlylysglylysasnclulclyleucllyValGlyAspG 483
1452 TACACTCATTTTGGATCAGACGAGGAGACGATTAAGCAAAACAGCCT 1501
483 ythrValilleucllysglnlysalaspsalasnbnlysalGlnAlap 500
1502 TTATGAAATCGGCTTGTCAGCGGACGGGTAGCGTCACTGAATGCC 1551
500 heserGlnValcllylevalserGlyArgserThrleuValleuasnsp 516
1552 GATTAATCATTCACCCGACCAACACTATTTGCGCTTTCGCGGAGAG 1601
517 asplysglnvalasprproasnserlethyrrheglynhearglylyar 533
1602 TTTGGATTTAAAGCGGCAATTCGCTTTCGTCACCGTATTCAAATACCG 1651
533 gleuaspleuasncllyasnserleuthrphasnhisleargasnIlea 550
1652 ATGAAGGGCGGATGATTCACACCAATCAAGACAAAGAAATCCAGCTT 1701
550 spaspelyalargyValValasnHisasmetThrAsnThrserasnle 566
1702 ACCATTACAGC.....ATAAAGATATTGCTACACCGG 1736
567 thrlethrcllygluserleuthrAsnProasnThrlethrSerty 583
1737 CAAT..... 1740
583 rsnilleglualaglnaspsasphisProleuarglleargserllep 600
1740 ..... 1740
600 rottyrarglnleuthrphasnnglnaspsasnargserlyrthrleu 616
1740 ..... 1740
617 lyslysglyalaserthrargsergileuProGlnasnserglyluse 633
1740 ..... 1740
633 rasngluasntrpleuthrmetelyargthrserasplalalysaya 650
1741 .....AACACAGCTTGATAGCAAAAGAAATTCCTTACACAGGTTGG 1785
650 snValmetasnhisleasnasncluarGmetasncllypheasncllytyr 666
1786 TTTGGCGAGAAAGATACGACCAAAAG...AACGGCGGCTCAACCTGT 1832
667 pheglyglulcluluthrlysalatrhnGlnasncllylysleuasnValth 683
1833 TTACAGGCCGCGGAGAAAGCCGACCTGCTGCTTTCGCGGAGACAA 1882
683 rphenasncllylyserasplelnasnargrheleuclulncllylthra 700
1883 ATTTAAACGGCAACATCACGCAAAACAAAGCAAACTGTTTTCACGGC 1932
700 snleuasnnglyasrpleuasnValGlnlysglythrleuPheleuSergly 716
1933 AGACCAACACCGGACGCTCATCATTCATTTAAAGACCATTCGCGGAAAA 1982
717 ArgProthrProhnsalatrGaspIlealaglylleserSerThrlysy 733
1983 AGAGGCG.....ATTCTCGCGGGGAAATCGTGGGACACAGCTGGA 2026
733 sasprProhnsrphenrclulasnasncllyValValGlnAspAsprpr 750
2027 TCAACCGCACATTTAAAGCGGAAACTTCCAATTTAAAGCGGACAGCGG 2076
750 leasnargasnphenlysalatrThrThmetasnValaThrGlyasnAlaser 766
```



```
2077 GTGGTTCC...CGCAATGTTGCCAAGTAAGCGATTCGATTGAG 2123
      :: ||| |||||::: |||::: ||
767 LeuTyrSerGlyArgAsnValAlaAsnIleThrSerAsnIleThrLase 783
2124 CAATCAGCCCAAGCAAGCTTTTGGTGTGCGACCGCATCAAGGCCACAA 2173
      |||||::: |||::: |||
783 rAsnAsnAlaGlnValHisIleGlyTyr.....LysThrGlyAspThrV 798
2174 TCTGTACAGCTTCGAGCTGAGCGGTGCAAAATGTGCGAAAACC 2223
      ::|||::: |||||::: |||::: |||::: |||::: |||
798 aLysValAlaTyrSerAspTyrThrGlyTyrValThrCysHisAsnSerAsn 814
2224 ATTACGACGATTAAGTATGCTTCATTCATAGACGACATCAGCG 2273
      ::|||::: |||||::: |||||::: |||||::: |||||
815 LeuSer...GluLysAlaLeuAsnSerPheAsnProThrAsnLeuArgI 830
2274 CAATGTGATCTTCGCGATCAGCTCATTTAAATCTCACAGGGCTTGCA 2323
      |||||::: |||||::: |||||::: |||||::: |||||
830 yAsnValAsnLeuThrGluAsnAlaSerPheThrLeu...GlyLysAla 846
2324 CACTCAACGCGCAATCTTAGTGCMAATGCGATACAGCTTATACAGTCAG 2373
      ::|||::: |||||::: |||||::: |||||::: |||||
846 snLeuPheGlyThrIleGlnSerIleGlyThrSerGlnValAsnLeuLys 862
2374 CACAAGCC.....ACCAAAAGCGCAACCTTAGCCTCGTGG 2411
      ::|||::: |||||::: |||||::: |||||::: |||||
863 GluAsnSerHisThrPheIleuThrGlyAsnSerAsnValAsnGlnLeuAs 879
2412 C.....AATGCCCA..... 2421
      ::|||::: |||||::: |||||::: |||||
879 nLeuThrAsnGlyHisIleHisLeuAsnAlaGlnAsnAlaAsnLysV 896
2422 ..GCAACATTAAATCAAGCCACATTAAAGCGCAACATCGGCTTCGGC 2469
      ::|||::: |||||::: |||||::: |||||
896 aLThrThrTyrAsnThrLeuThrValAsn.....SerLeuSerGly 909
2470 AATGCTTCATT.....AATCTAAGCGACA 2495
      |||||::: |||||::: |||||::: |||||
910 AsnGlySerPheTyrTyrValAspPheThrAsnAsnLysSerAsnLys 926
2496 GCGCGTACAAAC.....GCGACTGACGCT..... 2523
      ::|||::: |||||::: |||||::: |||||
926 sValValValAsnLysSerAlaThrGlyAsnPheThrLeuGlnValAla 943
2524 ..... 2526
      ::|||::: |||||::: |||||::: |||||
943 sPlyThrGlyLupProAsnHisAsnGlnLeuThrLeuPheAspLaser 959
2527 GGCACGCTAAGCAACGTAAGCCATTCCGCACTCAACGGTAATGCTC 2576
      ::|||::: |||||::: |||||::: |||||
960 AsnAlaThrArgAsnAsnLeuGlnValThrLeuAlaAsnGlySerVal.. 975
2577 CCTAGCCGTAAGCAAGTATTCATTGAAAGCAGCCGCTTACCGGAC 2626
      |||||::: |||||::: |||||::: |||||
976 .....AspArgGlyAlaTrpLysTyrLysLeuArgAsnValAsnGlyA 990
2627 AATCAGCGCGCGCAAGATACGCATTACACTTAAAGACAGCGAATGG 2676
      ::|||::: |||||::: |||||::: |||||
990 rGlyTyrAsp.....LeuTyrAsnProGlu... 997
2677 ACGCTGCGGTAGCGACGAAATTAGCAATTAAACCTTGACACGGCAC 2726
      ::|||::: |||||::: |||||::: |||||
998 .....ValGlnLysArgAsnGlnThrValAspThrThrAs 1009
2727 CATTCACACTATCCGCTATCCGACGATGCGGACGGCGCAAAACG 2776
      ::|||::: |||||::: |||||::: |||||
1009 nIleThrThrProAsnAspIleGlnAlaAspAlaProSerAlaGlnSerA 1026
2777 GCAGT..... 2781
      ::|||::: |||||::: |||||::: |||||
1026 snAsnGlnGlnIleAlaArgValGlnThrProValProProAlaPro 1042
2782 .....GCGACAGATGCGCGCGCGCGCTTCGCGCG 2813
```

```
1043 AlaThrGluSerAlaIleAlaSerGlnGlnProGluThrArgProAlaG 1059
2814 TTCGCGCGTTCCTATTATACGTTACA..... 2841
      ::|||::: |||||::: |||||::: |||||
1059 uThrAlaGlnProAlaMetGlnGluThrAsnThrAlaAsnSerThrGluT 1076
2841 ..... 2841
1076 hAlaProLysSerAspThrAlaThrGlnThrGluAsnProAsnSerGlu 1092
2842 .....CCGCCACTTC 2852
1093 SerValProSerGluThrThrGluLysValAlaGluAsnProProGlnI 1109
2853 GGTAGAACTC.....CGTTTCACACGCTGACGG 2881
      ::|||::: |||||::: |||||::: |||||
1109 uAsnGluThrValAlaLysAsnGlnGlnGluAlaThrGluThrProThr 1126
2882 TAAAGCGCAATTG.....ACGGT 2901
      ::|||::: |||||::: |||||::: |||||
1126 lAsnGlyGluValAlaLysGluAspGlnProThrValGluAlaAsnThr 1142
2902 CAGGAACATTCGCTTATGTCGGAACCTTCGGCTACCGCAGCGACAA 2951
      ::|||::: |||||::: |||||::: |||||
1143 GlnThrAsnGlnAlaThrGlnSerGlu.....GlyLysThrGlnI 1156
2952 ATTGAAGCTGCGGAAAGT..TCCGAAGCACTTACACCTTGGCGGTCA 2998
      ::|||::: |||||::: |||||::: |||||
1156 uThrGlnThrAlaGluThrLysSerGluProThrGluSerValThrValS 1173
2999 ACAATACCGGCAAGAACCT.....GCAACCTCGAACAA 3033
      ::|||::: |||||::: |||||::: |||||
1173 eArgL.....AsnGlnProGluLysThrValSerGlnSerThrGluAsp 1187
3034 TTGACGCTAGTGAAGGAAAGAC..... 3057
      ::|||::: |||||::: |||||::: |||||
1188 LysValValValGluLysGluLysAlaLysValGluThrGluGluTh 1204
3057 ..... 3057
1204 rGlnLysAlaProGlnValThrSerLysGluProProLysGlnAlaGluP 1221
3057 ..... 3057
1221 roAlaProGluGluValProThrAspThrAsnAlaGluAlaGlnAla 1237
3057 ..... 3057
1238 LeuGlnGlnThrGlnProThrValAlaAlaAlaGluThrThrSerP 1254
3058 ....AACAAACGCTGTCGCAAAACCTTAATTCACCTCAAAAGCAAC 3103
      ::|||::: |||||::: |||||::: |||||
1254 oAsnSerLysProAlaGluGlu..ThrGlnGlnProSerGluLysThrAsn 1270
3104 .....ACGTGATGCGGCGCTGCGGCTTACCAACATCATCGC 3141
      ::|||::: |||||::: |||||::: |||||
1271 AlaGluProValThrProValValSerGluAsnThrAlaThrGlnProTh 1287
3142 AAAGACGGCAGATTCCGCTGCATTAATCGGTCGCAAGACAGCTTTC 3191
      ::|||::: |||||::: |||||::: |||||
1287 rGluThrGlnGluThrAlaLysValGluLysGluLysThrGlnGluValP 1304
3192 CGACAAACCTGCGCAGCAGACGCAAAACGAGGCGGAAAGACACAGC 3241
      ::|||::: |||||::: |||||::: |||||
1304 roGlnValAlaSerGlnGluSerProLysGlnGluGlnProAlaAlaLys 1320
3242 CGCAAAACCTTGAACGCTGATTCGCGCGCGCGCATCGCTGCAAAAG 3290
      ::|||::: |||||::: |||||::: |||||
1321 ProGlnAlaGlnThr.....LysProGlnAlaGluProAlaArgI 1334
3291 .....ACAGAAAGCTTCGCAACCGCGCGCGCGCAAGCA.... 3324
      ::|||::: |||||::: |||||::: |||||
```

```

1334 uasnValLeuThrThrLysAsnValGlyGluProGlnProGlnAlaGlnP 1351
3324 ..... 3324
1351 roGlnThrGlnSerThrAlaValProThrThrGlyGluThrAlaAlaAsn 1367
3324 ..... 3324
1368 SerLysProAlaAlaLysProGlnAlaGlnAlaLysProGlnThrGluPr 1384
3325 ..GCGGGGAAATGTCGGCATTATGACGGCGAGAA..... 3360
::: |||||::: ::::::::::::::
1384 oAlaArgGluAsnValSerThrValAsnThrLysGluProGlnSerGlnT 1401
3360 ..... 3360
1401 hrSerAlaThrValSerThrGluGlnProAlaLysGluThrSerSerAsn 1417
3360 ..... 3360
1418 ValGluGlnProAlaProGluAsnSerLLeasnThrGlySerAlaThrTh 1434
3360 ..... 3360
1434 rMetThrGluThrAlaGluLysSerAspLysProGlnMetGluThrValT 1451
3360 ..... 3360
1451 hrGluAsnAspArgGlnProGluAlaAsnThrValAlaAspAsnSerVal 1467
3360 ..... 3360
1468 AlaAsnAsnSerGluSerSerGluSerSerArgArgArgSerVal 1484
3361 ...GAGAAAAACGGGTGACGGCGGATPAAAGACACCGCTTGGCGAAGC 3406
::: |||||::: ::::::::::::::
1484 lSerGlnProLysGluThrSerAlaGluGluThrThrValAlaSerThrG 1501
3407 AGCGGGAAGCGGAACCGGGCGGTACCGCCTTCCCGGGCGCGC 3456
||::: :::: |||||:::
1501 lngluThrThrValAspAsnSerValSerThrProLysProArgSerArg 1517
3457 CCGGCCCGCGGATTTGCCGCACTGCACCCCAACCGCAGCCCAACG 3506
||::: |||||::: ::::::::::::::
1518 ArgThrArgArgSerValGlnThrAsnSerThrGluProValGluLeuPr 1534
3507 GCAG..... 3510
1534 oThrGluAsnAlaGluAsnAlaGluAsnValGlnSerGlyAsnAsnValA 1551
3511 .....CGGACCTGATGACCGCTTATGCCAATGAC 3540
||::: |||||::: ::::::::::::::
1551 lAsnSerGlnProAlaLeuArgAsnLeuThrSerLysAsnThrAsnAla 1567
3541 GGTTCGATGGAATTTCCGCC.....AGCTCACAACGCT 3575
||::: |||||::: ::::::::::::::
1568 ValLeuSerAsnAlaMetAlaLysAlaGlnPheValAlaLeuAsnValG 1584
3576 TTTCCGCCGTCAGGACGATTAGCCGGCTA...TTTGGCGAAGCCGC 3622
||::: |||||::: ::::::::::::::
1584 yLysAlaValSerGlnHisLLeSerGlnLeuGluMetAsnGlnGlyG 1601
3623 GCAACGCCGTTTGACACAGCGCATCCGGACACCAACACTACCGTTGC 3672
::: |||||::: ||::: |||||:::
1601 lntyrAsnValTrpLieserAsnThrSerMetAsnLysAsnTrpSerSer 1617
3673 CAAGATTTCCGGCGCTACCGCCACAACCGACCTCCGCAATGGTAT 3722
::: |||||::: ::::::::::::::
1618 GluGlnTyrArgArgPheSerSerLysSerThrGlnThrGluLeuGlyTr 1634
3723 GCGAAGAAACCTCGCAGCGGG...CGCGTCGCATCTGTTTGGCAAC 3769
::: |||||::: ||::: |||||:::
1634 pAspGlnThrLLeSerAsnAsnValGlnLeuGlyGlyValPheThrTyrV 1651

```

```

3770 ACCGGACGAAAAACCTTCGACGAGCGCATCGCACTCGGACGGCTT 3819
||::: |||||::: ||::: |||||:::
1651 aLArgAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThr...Leu 1666
3820 GCCCAGCGCGCGTTTCGGGCATACGGCATCGAC...AGTTCATCAT 3866
||::: |||||::: ||::: |||||:::
1667 AlaGlnValAsnPheTrpSerLysTyrAlaAspAsnHisTrpTyrLe 1683
3867 CCGCATCAGCGCGCGGCTTTAGCGAGCGGACGCTTCAGACGGCA 3916
||::: |||||::: ||::: |||||:::
1683 uGlyLLeAspLeuGlyTyrGlyLysPheGlnSerAsnLeuGlnThrAsn 1700
3917 TCGGAGCAAAATCCGCCCGCGCTTCATTCAGCATTCAGCGACGA 3966
::: |||||::: ||::: |||||:::
1700 snAsnAlaLysPheAlaArgHisThrAlaGlnLeuGlyLeuThrAlaGly 1716
3967 TACCGCGCGGTTTCGGCGATTTCGCATCGAACCGCACATCGCGCAAC 4016
::: |||||::: ||::: |||||:::
1717 LysAlaPheAsnLeuGlyAsnPheAlaValLysProThrValGlyValAr 1733
4017 GCGCTATTGCTCCAAAGCGGATTAACCGCTACGAAACGTCAATATCG 4066
::: |||||::: ||::: |||||:::
1733 gTyrSerTyrLeuSerAsnAlaAspPheAlaLeuAlaGlnAspArgLLe 1750
4067 CCACCCCGCGCTTCATTCACCGCTACCGCGCGCATTAAGCGCAT 4116
::: |||||::: ||::: |||||:::
1750 yValAsnProLLeSerValLysThrAlaPheAlaGlnValAspLeuSer 1766
4117 TATTCATTCACCGCGCGCACACATTCATTCACCGCTTATTCAGCT 4166
||::: |||||::: ||::: |||||:::
1767 TyrThrTyrHisLeuGlyGlu...PheSerLLeThrProLLeuSerAl 1782
4167 GTCTTATTCAGCATCGCGCTTCGGCAAGTCCGACACCGCTCAATATCG 4216
||::: |||||::: ||::: |||||:::
1782 aArgTyr...AspAlaAsnGlnGlyAsnGlyLysLLeAsnValSerValT 1798
4217 CCGTATTCGCTCAGATTTCCGCAAAACCGCATCGGAAATGGCGCTA 4266
::: |||||::: ||::: |||||:::
1798 yAspPheAlaTyrAsnValGluAsnGlnGlnGlnThrAsnAlaGlyLeu 1814
4267 AACGCCCAATCAAGTTTCACGCTGCCCTCGACGCGCGCGCCCAA 4316
||::: |||||::: ||::: |||||:::
1815 LysLeuLysTyrHisAsnValLysLeuSerLLeLLeGlyLeuThrLry 1831
4317 AGGCCCGCAACTGGAAGCGCAACACAGCGCGGCATCAATTAAGCTAC 4365
||::: |||||::: ||::: |||||:::
1831 sAlaLysGlnAlaGluLysGlnLysThrAlaGluValLysLeuSerPhe 1847

seq_name: SwissProt_40:YPIA_ECOLI
seq_documentation_block:
ID YPIA_ECOLI STANDARD; PRT; 1569 AA.
AC P52143; P76610; P77017; P77019;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical outer membrane protein YPIA.
GN YPIA OR B2647.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).

```

```

196 eleuileasprcllygcyserglnllevalylsvalglnlglnhisalasp6 213
      :||| ||||| |||
      :||| ||||| |||
213 1ythrthrlleasnlyserclyserglnsaprvalglnlglyserleu 229
      :||| ||||| |||
      :||| ||||| |||
832 AAGCAAAAGTGTTAAATTAAATGGGCTATTCGAACGGCAACCCCTATAT 881
230 Alathrsntrhtrlleasnly.....Glyarglnthyva 242
882 AGGA.....AAAGCAATGGCTTCGAC 904
242 lgluInserthrvalgluThrtrhtrllelysasnlglyglnclnA 259
905 TGCTTCGTAAGATGCTGTCTATGATGCAAAATCTTCTCGACATCCAT 954
259 rGValTygluSerargAlaLeuaspThrtrhtrllegluInglyThrln 275
955 TCAGATATTCAGCAACACGTCGCAAAATGGCAATTACTCTTTAAGCAC.. 1007
276 SerleuasnSerlySerThrAlaLysAsnThrhislleTyrSerclYgl 292
1003 .....GATATATATGCGACAGAGAAATATCAATGCGCAAAACAG 1033
292 yThrGlnllevalAsprsnrthrserhsaprvalllegluValTys 309
1040 AACCAATTCCTCGCTCAATAGA..... 1067
309 erglGlyValleuAsprValarglYglYThrAlathrsanValThrln 325
1063 .....TTAAAAACAGCAACCGCTCAATGTTTAATGTTCT 1097
326 HisaprlYAlalleuulysThrAsnThr.....AsnlglyThrth 339
1098 TTTATCCGACACAGCAAGAACCTTTAT.....CATGCG 1135
339 rValSerGlyThrAsnSerGluGlyAlarheserlleHisasnHisvalA 356
1136 CA.....GGTGGT.....GTCAACAGTTAT 1155
356 laAsprsnValleuLeuGluAsnGlylnHisleuAsprlleasnrlayr 372
1155 ..... 1155
373 GlySerAlaasnlysrhrlleleuysasprlyglYhrMetSerValle 389
1156 .....CGAACCGAGTGAATATAGA..... 1176
389 uThrAsnAlaLysAlaAsprAlaThrArglleAspsnclYglValMetA 406
1177 .....GAAAT 1182
406 sprValAlaGlYasnAlaThrAsnThrllleleasnclYglYThrGlnsn 422
1183 ATTTCCTTATATGACAGCAAGAAAGCGCAATGTACTTACCAACAAT 1232
423 lleasnAsnTyglYlleAlaThrGly.....ThrAsnll 434
1233 CAATCAAGGT.....GCTGAGAGATTAATTTCCAAAGAG 1267
434 eAsnSerGlyThrGlnAsnllleYserclYglY.....LysAlaA 448
1268 ATTTTACGGCTCGCTGAATATAC.....GAAACTGGG 1302
448 sphrThrTrlleleaserclYserArglnValAlaGluValAsprclY 464
1303 CAAGCGCGGCGCTCATATCACTAGTGAAGACAGTACCGTACTTGGAAAGT 1352
465 ThrAlaTrlleglYserAsnllleSerAlaGlylYserleuileValTythr 481
1353 AAACGGCGTGGCAACGAC.....CGCGTGT 1378
      :||| ||||| |||
      :||| ||||| |||

```

481 rgllyglilealnlhlsglyvalasnlglnluthrglyseralaleuvala 498
1379 CCAAAATCGCGCAAGCAGC.....CTGCAC 1404
498 laasnhrhrglylaglythrapsllegllytrhsnlyseuserhs 514
1405 GTTCACGCCAAGG.....GAAACCAAG 1430
515 Phehrthlethrhlrglygluaalaasnlyrvalvalleuglnluthrgl 531
1431 CTGATCGAGGTGGCGAGCGTACAGTATTGGATTCACAGCAGCAGC 1480
531 yglulethrvalvalalalysthrseralalysasnthrthrlleaspr 548
1481 ATAAAGCGCAA.....AAACACCTTTAGTGAATCGGCTTG 1518
548 hrhlygllyslleuilevalglnglualalysthrapsrthrarg 564
1519 GTACGCGGAGGGGTACGTCACATGATCCGAT..... 1554
565 leuasnasnlygllyvalleuvalglualaspolygllylualalysnl 581
1555AATCAGTCAACCCCGCAAACTCTATTGCGCTTCGCGGCGAC 1600
581 svalglnglnlserlgllyalaleuilealaserthrthrsrclgt 598
1601 GTTTG.....GATTTAAGCGGCGATTCGCTTCACCGTATTCAA 1644
598 hrleuilegluglythraserlgllyaspalaphetyr...learg 613
1645 AATCCGATGAGGGGATGATGTCACACCAATCAAGCAAAAGATC 1694
614 asnsrvalalalysasnvalleu.....glusnlaglyse 627
1695 CACCGTTACATTACAGGCATAAAGATTTGTACACCGCAGATTAACA 1744
627 rleuthrvalvalthrlyserargalalaspthrilleasnlaa 644
1745 ACAGCTTGATAGCAAAAGAAATTCGCTCAACGTTGGTTGGGAG 1794
644 sn.....gllysmetaspal.....tyrly... 651
1795 AAAGAT.....ACGACCAAAAGCAGCGCGCTCAACCTTTGTTACA 1838
652 lysasprvalgllythrvalleuasnserlagllythrlnthrletryal 668
1839 GCCCGCCGAGAGACCGCAGCTGCTCTTCGCGGCAACAATTAA 1888
668 aseralathrserasplysalaasnillelysgllylsglnthrvalt 685
1889 ACGGCATCATCAGCAACCAAC.....GGCAACTGTTTTTCAGC 1929
685 yrglyleualathrglualasnillegluserlgllyglnglnllevalasp 701
1930 GCGAGACCAACACCGCAGCCTACATCTTAAACGAC..... 1968
702 gllyllyserthrglulys.....Thrhlslleasnlygllythrlnth 716
1969C 1969
716 rvalglhlnsnlyrglyllyslalaileasnthrapslilevalserllyeug 733
1970 ATTGGTCGCAAAAGAGGCGATTCCTCGGAGGAATCGTG...TGGAC 2016
733 lnglnlleuetaalasnlythrhlalagllyserlilleleasnlyly 749
2017 AACGACTGATCAACCGCATTTAAAGCGGAAACTTCCAAATTAAAG 2066
750 serglvalvalasnlygllyleualaglualasnservalleuasn 766
2067 CGGACAGCGGTGTTCCTCCGCAATGTCGCAAAAGTGAAGCGGATTGGC 2116
766 pellygllythrleuaspvalargglulyselyseralathrgly..... 780

2117 ATTTCAGCATCAGCCCAACAGCATTGTTGGTGCGACCGCATCAAGC 2166
781 ..lleglnlserseglnglylaleuvalalathrthrargalathr 796
2167 CACACATCTGTACACGTTGCGACTGAGCGGCTGTGCAAAATGTGTGCA 2216
797 argvalthrghlythrargalasp.....glyalalapheserilegl 811
2217 AAAA.....ACA 2224
811 uglnglylalaalasnalleleuleuvalaasnlygllyvalleuthrv 828
2225 TTACGAGAT..... 2235
828 algluseraspthrseraspyslthrnglnvalasnmetclglyarg 844
2236AAGTATTGCTTCATTGACTAAGACCGACATCAGCGCAA 2276
845 glullevallyslthrlyalathrhlthrhlthrleuthrhlgl 861
2277 TGTGATCTT.....GCCATCAGCGCTCATTTAATCTCACAG 2314
861 yglnglnllevalgllylvalalasnlglnluthrthrlleasnspglyc 878
2315 GCGTGCACACTCAACGCGCATCTTAGTCAATGGCAT..... 2355
878 lytleglthr.....valseralalasnlyglulathlelys 890
2356 ACACGTTATACAGTCAGCCACACCGCCCAAAAGCGCAC..... 2397
891 thrlylileasnlygllylthrleuthrvalalasnspasnlyllyal 907
2398 ..CTTACCTCGTGGCGAATGCCAGCAACATTATATCAACGCATTA 2446
907 athrasplilevalglasnserclyalalaleuenglnthrserhrralaa 924
2447 ACGCAACACATCGGCTTCGCGCAT.....GCTTCATTATCTA 2487
924 snly...lleglnlserlgllythrhlslngllythrpheserile 939
2488 ACGGACACCGCGTACAAAAGCGAGCTGACGCTTCGCGCACTAA 2537
940 sergllyasnleualathrasmnleuleuaglualasnlygllyasnleu 956
2538 GCGAAACGTAAAGCATTCGCACTCAACGTAATGTCTCCAGCGATA 2587
956 uvalleualagllythrhlalargaspserthrvalgllylly 972
2588 AGCGAGTATTCATTTGAAGCAGCGCTTACCGGCAAAATCGCGGC 2637
973 ..Alametglasnleuglylgnaspseralathrlyslalasnsergly 988
2638 GCGAGGATTCGCGCATTTACCTTAAAGACAGCGAATGAGCCTGCCGTC 2687
989 llyglnglnlthrleucllyargserlyspasgluphaglinalaleuana 1005
2688 AGCGACGGAATTA..... 2700
1005 glalgluaspleuglnvalalagllythrhalalevaltryalaglyt 1022
2701GGCAATTAAACCTT... 2715
1022 hrleualaspalaserlaserlglalathrlyserleuserlsermet 1038
2716GACACGCGCAC...ATTACATCAATTCGCGCTATGCG... 2751
1039 Thrprargaspasnvalthrprovallylbleugllylvalalargtl 1055
2752 ..CAGATCGCGGAGCGGCGCAACCGGCGACTGCGACAGATCGCGCGCC 2800
1055 ethraspseralathrleuthrleuglyasnlylvalaspthrthleua 1072

```
2801 GCCGTTCCGCCGCTTCCGCCCT..... 2823
      ::::|
1072 laAspLeuThrAlaIaSerArgIysValTrpLeuAsnSerAsn 1088
2824 .....TCCCTATTAA.. 2832
      |||||
1089 SerCysAlaGlyThrSerAsnCysGluTyrArgValAsnSerLeuLeu 1105
2833 .....TCCGTTACACCGCCACTCCG 2854
      ::::|
1105 uAsnAspGlyAspValTyrLeuSerAlaGlnThrAlaIaProIaIaThr 1122
2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAATTGAACGGTAC 2904
      ::::|
1122 hrAsnGlyIleTyrAsnThrLeuThrAsn...GluLeuSerGlySer 1137
2905 GGACATTTCCGCTTATGTCGGAACCTTTGGCTACCGCAGCGCAAAATT 2954
      |||||
1138 GlYsnPheTyrLeuHisThrAsnValAlaGlySerArgIysAspGlnLe 1154
2955 GAACGTTGGCGAAAGTTCCGAAGCACTTACACCTTGGCGGTCAACATA 3004
      ::::|
1154 uValValAsnAsnAsnAlaThrGlyAsnPheIysIlePheValGlnAsp 1171
3005 CCGGCAACGAACTCGCAAGCTCGAACATTTGACGGTAGGAAGAAA 3054
      |||||
1171 hrGlyValSerProGlnSerAspAspAlaMetThrLeuVal..... 1184
3055 GACACAAACCGCTGCCGAAACCTTAATTCACCTCGCAAAACGAA.. 3102
      |||
1185 .....LysThrGlyGlyGlyAspAlaSerPheThrLeuGlyAsnThrG 1199
3103 ....CACGTGATCCGCGCGCTGACCACTCATCCGCAAGACG 3148
      |||||
1199 yGlyPheValAspLeuGlyThrTyrGluTyr...ValLeuIysSerAsp 1215
3149 GCGAGTCCGCGCTGCATTAATCCGCTCAAGAACAGACTTCCGACAAA 3198
      |||
1215 Ly.AsnSer..... 1217
3199 CTCGGCAAGGCAAGCAAAACCAAAACAGCGGAAAGAACACGCGCAAG 3248
1218 .....AsnTTPA 1220
3249 CCTTGACGGCTGATTGCGCGCGCGCATGCCGTGCAAAAGACAGAAA 3298
      |||||
1220 snLeuThrAsnAspValIysProAsnProAspPro..... 1231
3299 GCGTTCGCAACGCGCGCGGCAAGCGGGGAAATTCGGCATTTATG 3348
      ::::|
1232 ...IleProAsn..... 1234
3349 CAGGCGGAGAAAGAAAAACGGGTGACGGGATAAAGACACCGCTT 3398
1234 ..... 1234
3399 GCGGAACACAGCGGAAGGAAACCGCGGCTACCAACCGCTTCCCC 3448
1235 .....ProIysProAspProIysProAspProIysPro 1245
3449 GCGCCCGCGCGCGCGCGGATTTGCGCAATGCAACCCACACGAG 3498
      |||
1246 AspPro.AsnProIysProAsp...ProThrProAspProThrProThr 1260
3499 CCGCAACGCAACGCGACCTGATCAGCGGTATGCAATAGCGGTTTATG 3548
      |||
1261 ProValProGluIulYsArgIleThrPro.....SerThrAlaIaValLe 1275
3549 TGAATTTTCCGCGCTCAACAGCGTTTTC.....GGCG 3583
      ::::|
1275 uAsnMetAlaIaIaThrLeuProLeuValAlaPheAspAlaIuLeuAsnSer 1292
3584 TACAGACGAATTTAGACCGCTATTGCGCAAGACCGCGCAACGCGCT 3633
```

```
1292 leArgIuThrLeuAsnIleMetLysAlaSerProHisAsnAsnVal 1308
      ::::|
3634 TGCAGACAGCGCATCCGGACACCAACACTACCGTTCCGCAAGATTTCG 3683
      |||
1309 Trp...GlyAlaThrTyrAsnThrArgAsnAsnValThrThsAlaG 1324
3684 CGCTACCGCCCAACAAACCGACCTGGCCAAATCGTATGCAAAAAAC 3733
      |||
1324 yAlaGlyPheGluGlnThr.....LeuThrGlyMetThrValGlyI 1338
3734 TC.....GGCAGCGCGCGGTCCGCTCTG 3759
1338 leAspSerArgAsnAspIleProGluGlyIleThrThrLeuGlyAlaPhe 1354
3760 TTTTCGACACACGCGACCGCAAAACACTTCGACAGCGC..... 3798
      ::::|
1355 MetGlyTyrSerHisSerHisIleGlyPheAspArgIysGlyHisGly 1371
3799 .....ATCGGCACTCGGACAGCGCTTCCACAGCGCG 3832
1371 rValGlySerTyrSerLeuGlyGlyTyrAlaSerTrpGluHisGluSer 1388
3833 TTTTC.....GGCAATACGGCATCGACAGGTC.....TAC 3864
1388 LysPheTyrLeuAspGlyValValIysLeuAsnArgPheLysSerAsnVal 1404
3865 ATCGCAATCAGCGCGCGCGGCTTTAGACAGCGCACCTT...TCGA 3911
      |||
1405 AlaGlyLysMetSerSerGlyGlyAlaAlaIysSerTyrHisSerAs 1421
3912 GGGCATCGGAGGCAAAATCCGCGCGCGCTGCTGCATTAACGCAATTCAG 3961
      ::::|
1421 nGlyLeuGlyGlyHis.....IleGluT 1429
3962 CACGATACCGCGCGCTTCCGGGATTCGGCATCGACCGCACATCGGC 4011
      ::|
1429 hrGlyMetArgPheThrAspGlyAsnTrpAsnLeuThrProTyrAlaSer 1445
4012 GCACGCGCTATTTCGTCCAAAAAGCGATTACGCTAGCAAAAC..... 4056
      |||
1446 LeuThrGlyPheThrAlaAspAsnProGluTyrHisLeuSerAsnGly 1462
4057 .....GTCAATATCGCCACCGCGCTTGCAATTCACCGCTACCGC. 4098
1462 LysSerIysSerValAspThrArgSerIle.....TyrArg 1475
4099 ..GCGGCAATTAAGCAGATTAATTCATTCAAACCGGCAACACATTCC 4146
      |||
1475 IuLeuGlyAlaThrLeuSerTyrAsnMetArgLeuGlyAsnGlyMetGlu 1491
4147 ATCAAGCCTTATTGAGCGCTGCTCATACGATCCGCTTCGGCAAAAGT 4196
      ::|
1492 ValGluProIleThrLeuLysAlaAlaValArgLysIuPheValAspAs 1508
4197 CCGAACAACGCTCAATACCGCGGTATGCTCAGAGATTTC.....GGCA 4240
      |||
1508 nArgValIysValAsnSerAspGlyAsnPheValAsnTyrLeuSerGly 1525
4241 AAACCCGAGTGGCGAATGGGCGCTAAACGCCGAATCAAGGTTTCAG 4290
      ::|
1525 rGArgGlyIleTyrGlnAlaGlyIleLysAlaSerPheSerSer...Thr 1540
4291 CTGTCCCTCCACGCTCGCGCGCGCAAGGCGCGCACTGGAACGGCAACA 4340
      |||||
1541 LeuSerIlyHisLeuGlyValGly.....TyrSerHis 1551
4341 CAGCGCGGCAATCAA.....TTAGCTTACCGCTGG 4371
      ::::|
1551 sSerAlaGlyValGluSerProTrpAsnAlaValAlaGlyValAsnTrp 1567
```

seq_name: SwissProt_40:PERT_BORBR


```
279 GlyProLeuLeuAspGlyTyrTyrGlyVal..... 288
2223 CATTACGACGATAAAGTATGCTTCATTGACTAGACCGACATCAGC. 2271
289 .....AspValSerA 292
2272 ..GGCAATGTCATCTTGGCGATCAGCTCATTTAAATCTCACAGGGCTT 2319
292 sPserThrValAspLeuAlaGln.....SerLeuValGluAla 304
2320 GCCACACTACAGCGCAATCTTAAGCAATGGCGATACACGTTATACGT 2369
305 ProGlnLeuGlyAlaAlaIleArgAlaGlyArgGlyValIleThrVal 321
2370 CACCCACACAC.....GCCACCCAAAAGCGCAACCTTACCTCGCGGGA 2413
321 lSerGlyGlySerLeuSerIleAlaProHisGlySerValIleGluThrGlyG 338
2414 ATGCCCAACGCACTTTAAT.....CAAGCCACATTA 2445
338 lYcIlyAlaIleArgArgPheProProIleAlaSerProLeuSerIleThrIleu 354
2446 AACGGCAACATCGGCTTCGGCAATGCT...TCATTTAATCTAAGCA 2492
355 GlnAlaGlyAlaIleArgAlaGlnGlyArgAlaLeuIleuTyrArgValLeuPr 371
2493 CCACGGCTACAAACGCGAGTGTGACGCTTCGCGCAACGCTAAGGCA 2542
371 oGluProValLys.....LeuThrLeuAlaGlyGlyAlaGlnGly. 384
2543 ACGTAAACCATTCGCGACTCAACGGTAAATGCTCCCTACCGATTAAGCA 2592
385 .....GlnGlyAspIleValAlaIleThr..... 391
2593 GATTTCATTGTAAGACGCGCTTACCGGCAACATCAGCGCGGCA 2642
392 .....GluLeuProProIleProGlyAlaSerSerGlyProle 404
2643 GGATACGATTACACTTAAGAAAGACGCAATGACGCTGCGGTAGCA 2692
404 uAspValAlaIleu...AlaSerGlnAlaIleArgTyrPhe...GlyAlaIleThr 419
2693 CGCAATTACGCAATTTAAACCTTGACACACGCCATTAACCTCAATTC 2742
419 rGlyAlaValAspSerLeuSerIleAspAsnAlaIleThrIleValMet..... 433
2743 GCCTATCGCAGATGCGGCGACAGGGCGCAACCGGACGTGCGACGATGC 2792
434 .....ThrAsp... 435
2793 GCGCGCGCGCGCTTCGCGCGCTTCGCCGCTTCCCTATTATCCGTTAC 2842
436 .....AsnSerAsnValGlyAlaIleLeuArgLeuAlaAs 446
2843 CGCCAACTTCGCTAGAA.....TCCCGTTTCAAC 2871
446 eAspGlySerValAspPheGlnGlnProAlaGluAlaGlyArgPheLys 462
2872 ACGCTGACGTTAAACGCAATTAAGAGGTGACGGAACATCCGCTTAT 2921
463 CysLeuMetValAsp...ThrLeuAlaGlySerGlyLeuPheArgMetAs 478
2922 GTCCGAACCTTTCGGCTACCGGACGCAAAATTTGAAGCTGGCGGAAGTT 2971
478 nValIleAlaAspLeuGlyLeuSerAspLysLeuValValMetArgAspA 495
2972 CGGAAGGACACTTACACCTTGGCGGTACCAATACCGGACGACGACTGCA 3021
495 lAspGlyGlnHisArgLeuLeuValArgAsnSerGlySerIleProAla 511
3022 ACCCTCGAACAATTCAGCTAGTAGAAGAAAGACAAACAAACCGCTGTC 3071
512 SerGlyAsnThrMetLeuLeuValGln.....ThrProArgGly 524
3072 CGAAACCTTAATTTCACCTGCAAAACGAA.....CACGTGCATGCCG 3115
524 ySerAlaIleThrPheThrIleuAlaAsnLysAspGlyLysValAspIleG 541
3116 GCGCGTGGCGTTTACCACTCATCCGCAAAAGAGGCGAGTTCCGCTGCAT 3165
541 lYThrTyrArgTyrArgLeu..... 547
3166 AATCCGTTCAAAAGAACAGAGCTTTCGACAACTCGGCAAGCGCAAGC 3215
547 ..... 547
3216 CAAAAACAGCGGCAAAAGACAAACGCGCAACGCTTGACGCTGATTG 3265
548 .....AlaAlaAsnGlyAsnGlyGlnTyr.....SerLeuValG 559
3266 CGCGCGGCGCGATGCCGTGCAAAAAGACAGAAACGTTGCCGACCGGCG 3315
559 lYAlaIleLysAlaIleProIleAlaPro.....LysPr 568
3316 CGCGACGCGAGCGGGGAAATGTCGCATTATGACAGCGGAGAAAGAA 3365
568 oAlaProGlnProGlyPro..... 574
3366 AAAACGGTGCAGCGCGATTAAAGACACCGCTTGCGCAACAGCGCGAAG 3415
574 ..... 574
3416 CGAAACCCGCGCGGCTTACACACCGCTTCCCGCGCGCGCGCGCGCG 3465
575 ..GlnProGlyProGlnProIleProGln.ProIleProGlnProIlePro 590
3466 CGGATTTTCCCGCAACTGCAACCCCAACCGGACCGCAACCGCGCGA 3515
590 roGlnArgGlnProGluAlaIleProAlaProGlnProIleProAlaIleArgGly 606
3516 CCTGATACGCGCTTATGCAATACCGGTTGAGTGATTTTCCCGCCACG 3565
606 uLeu...SerAlaAlaAlaAsnAlaAlaValAsnThrGlyGlyValGlyL 622
3566 TCAACAGCGTT.....TTCGCGCTACAGACGCAATTAACCGCGCTTAT 3609
622 euAlaSerThrLeuTrpTyrAlaGlnIleSerAsnAlaIleuSerLysArgLeu 638
3610 GCCCAAGACCGCGCCAC.....GCCGTTTGCAAAACGCGCAT 3647
639 GlyIleuLeuArgLeuAsnProAlaSpAlaGlyGlyAlaIleArgGlyPhe 655
3648 CGCGGACACCAACACTACCGTTTCGCAAGATTTCCGCGCTACCGCGCAC 3697
655 eAlaGlnArgGlnGlnIleuAspAsnAlaGlyAlaGlyArgArgPheAspIle 672
3698 AAACCGACTGCGCCCAATCGGTATGCAAAACCTC..... 3735
672 yValAlaIleGlyPheGlnIleuGlyAlaAspHisAlaValAlaValAlaGly 688
3736 GCGACGCGCGCGCTGCGCATCTGTTTCCACACACGCGGCAAAACAC 3785
689 GlyArgTyrPheIleuGlyGlyLeuAlaGlyTyrThrArgGlyAspArgGly 705
3786 CTTC...GACGACGCGCATCGCAACTCGGACGCGCTGCCGACGCGCGCG 3832
705 yPheThrGlyAspGlyGlyGlyIleHisThrAspSerValHisValGlyGly 722
3833 TTTTTCGCGCAATTCGCGATGCAAGGTTCTTACATCGGCATACGCGGCGG 3882
722 yAlaIleThrTyrIleAlaAsnSerGlyPheTyrIleuAspAlaIleThrLeuArg 738
3883 GCGGCTTTTACGACGCGGACGCTT.....TCGACGCG..... 3915
739 AlaSerArgLeuGlnAsnAspPheLysValAlaGlySerAspIlyTyrAl 755
```


130ValSerAspThrIArg 135
1808 AACGACGGGCGGCTACACTTGTTACCACCCGCC 1845
135 sPaSPaSPgLYlLeAlaLeuTYrValAlaIglYglInAlaIglInAlaSer 151
1846 GCAGAGACCGCACCTGCTGCTTTCCGGGGAACAATTTAAA 1889
152 lLeAlaSPSerThrLeuInglYAlaIglYlYAlaIrgValaIgluaIrgl 168
1890 CGGACATCAGCGCAACAAACGCAACTGTTTTCAGCGGC 1932
168 yAlaSPValThrValIglInaIrgSerThrIleValaSPglYglYLeuHisI 185
1933AGACCAACACCGCAGCCCTACATCATTTAAAGCACCAT 1971
185 lglYThrLeuGlnProLeu 191
1972 TGGTCGCAAAAAGAGGCAATTCCTCGGGGAAATCGTGGGACAACGA 2021
192GlnProGlnaSPLeuProProSerArgValaValLeuGlnaSPTh 206
2022 CTGGATCAACGCACAT 2038
206 rSerValThrAlaValaProAlaSerGlYAlaProAlaAlaValaPheValP 223
2039 TTAAGCGGAAACTTCACAAATTAAAGCGGACAGCGGCTGTTCCCGC 2088
223 heGlYAlaSPInglYLeuThrValaSPglYglYHisIleThrIrglYglYArg 229
2089 AATGTTGCCAAAGTGAAGAGCGATTGGCAATTGACAAATCAGCCCAAGC 2138
240 AlAlaIleLYValAlaAl 245
2139 AGTTTGTGTGTGCACCGCATCAAGCCACACATCTGTACCGTTCCG 2188
245 aMetAspGlYAlaIleValaHisLeuclInaIrgaIaThrIleArgIrgLYa 262
2189 AC 2190
262 sPaIaProAlaIglYlYAlaValaProglYglYAlaValaProglYlYAla 278
2191TGGACGGGTCTGCAAA 2207
279 ValProglYglYpHeGlYProLeuLeuAspGlYTrpTYrGlYAla 293
2208 TTGTGTGAAAAAACCATTAACGAGCATTAAGTATTGCTTCATTGACTA 2257
293 293
2258 AGACGACATTCAGC...GGCAATGTGCATCTGGCGATCAGCGTCATTTA 2304
294AspAlaSerAspSerThrValaSPLeuAlaIglIn 304
2305 AATCTCACAGGCGCTTCCACACTCAACGCGCAATCTTAGTCAAAATGGCGA 2354
305 SerIleValaIgluaIaProglInLeuGlYAlaIleArgAlaIglYArgI 321
2355 TACAGCTTTACAGTCAGCCACAC...GCCACCCAAAACGGCAAC 2398
321 yAlaIrgValaThrValaSerGlYglYSerLeuSerAlaProHisIglYanY 338
2399 TTAGCCTGTGGCAATGCCCAAGCAACATTTAAT 2433
338 alIleGlYThrIrglYglYAlaIrgaIrgPheProProAlaSerPro 354
2434 ...CAAGCCATTAAGCGCAACATCGGCTTGGCAATGCT...TC 2477
355 LeuSerIleThrLeuGlnaIglYAlaIrgaIrgaIrgInglYArgAlaLeuLe 371
2478 ATTATATCAAGCGACGCGGTACAAAGGCGAGCTGACGCTTTCCG 2527
371 uTYrATyValLeuProGlnProValIysLeuThrLeuAlaIag 385
2528 GCACGCTTAAGCGCAACGTAAGCCATTCCGACCTCAACGTAATGTCTCC 2577
385 lYglYAlaIrgInglYGlnIYAspIleVal 394
2578 CTAGCCGATTAAGGAGTATTCATTTTGAACGACCGCTTTACGGACA 2627
395 AlAThrGluLeuProIleProglYAl 404
2628 AATCAGCGCGCAAGATACGCATTAACCTTAATAAGACACGCAATGA 2677
404 aSerSerGlYProLeuAspValaIaLeu...AlaSerGlnAlaIrgTrp 420
2678 CGCTGCGTCAGCAGCAGGAATTAGCAATTTAACCTTGACACCGCAC 2727
420 hr...GlYAlaThrArgAlaValaSPSerLeuSerIleAspAsnIaThr 435
2728 ATTACACTCAATTCGCCCTATGCGCACGATGCGGAGGCGCAACCGG 2777
436 TrpValMet 438
2778 CAGTGACAGATGCGCGCGCGCGCTTCGCGCGCTTCGCGCGTCC 2827
439ThrAspAsnSerAsnValaIrgY 446
2828 TATTATCCGTTACACCGCACTTGCTAGAA 2859
446 lAlaIrgLeuAlaSerAspGlYSerValaSPpHeGlInProAlaIglu 462
2860 ...TCCCGTTTCACACGCTGACGTTAAAGCGCAATTGAACGGTCACG 2906
463 AlagIYArgPheLYsValLeuMetAlaSP...ThrLeuAlaIglYSerGl 478
2907 AACATTCGCTTTATGTGGAATCTTTCGCTACGCGACGCAAAATGA 2956
478 yLeuPheArgMetAsnValaPheAlaAspLeuGlYLeuSerAspLYsLeuV 495
2957 AGCTGCGGAAAGTTCGAGAGCACTTACACCTTTCGCGCTGCAACAATAC 3006
495 alValaMetArgaSPAlaSerIglYglInHisArgLeuTrpValaIrgaSPSer 511
3007 GCGAAGCACTGCAAGCCTGCAACAAATTAAGCGTAGTGAAGAAAAA 3056
512 GlYSerGlnProAlaSerGlYAsnThrMetLeuLeuValGln 525
3057 CAACAACCGCTGTCCGAAACCTTAATTACCTGCAAAACGA 3102
526 ...ThrProArgGlYSerAlaIaIaThrPheThrLeuAlaSPAsp 541
3103 ...CACGTGATGCGCGCGCTGGGCTTACCAACATCCGCAAGAGAGGC 3150
541 lYlYsValaSPleGlYThrTYrArgTYrArgLeu 552
3151 GAGTTCGCGCTGCATTAATCCGTCAAAGAACAGAGCTTCCGACAAACT 3200
552 552
3201 CGGCAAGGCAACCAAAAAACAGGGGGAAGAAAGACAGCGCAAGCC 3250
553AlaAlaAspGlYAsnGlYAlaIrgTrp 560
3251 TTGACGGCTGATTTGCGCGCGCGCGATCCGCGAAAGACAGAAAGC 3300
561 ...SerLeuValaIglYAla...LYsAlaProProAlaProLYsPr 573
3301 GTTCCGACACCGCGCGCGAGCAGCGCGGGAATGTCGCAATTATGCA 3350
573 oAlaPro...GlnProglYProGlnProglYProGln 584
3351 GCGGAGAGAAAGAAAAACGGGTGCGAGGCGATTAAGACACGCGCTTG 3400
585ProPro... 586


```
2680 CTGCGCTGAGCGCAGCAATTAGCAATTTAAACCTTGACAAAGCCACCAT 2729
      ::::: ||| ||::: ::::: |||
425 ...GlyAlaThrArgAlaValAspSerLeuSerIleAspAsnAlaThr 440
      ::::: ||| ||::: ::::: |||
2730 TACACTCAATTCGGCTTATCGCCACGATGCGGAGGGCGCAACCGGCA 2779
      ::::: ||| ||::: ::::: |||
440 pValMet..... 442
2780 GTGCACAGATGCGCGCGCGCGCTTCGCGCGCTTCGCGGTTCCCTA 2829
      |||||
443 ...ThiAsp.....AsnSerAsnValGlyAla 450
2830 TTATCCGTTTACACCGCCAACTTCGCTAGAA..... 2859
      ||| ::::: |||||
451 LeuArgLeuAlaSerAspGlySerValAspPheGlnGlnProAlaGluAl 467
2860 .TCCCGTTTCAACAGCTGAGCGTTAAAGCGCAATTGACGTCAGGAA 2908
      ::::: ||| ||::: ::::: |||
467 aglyArgPheLysValLeuThrValAsn...ThrLeuAlaGlySerGlyL 483
2909 CATTCGCGCTTATGTCGGAACCTCTCGGCTACCGCAGCAACAATTGAA 2958
      |||||
483 euphArgMetAsnValPheAlaAspLeuGlyLeuSerAspLysLeuVal 499
2959 CTGCGGGAAGTTCGGAAGCACTTACACCTTGGCGGTCAACAATATCCGG 3008
      ::::: ||| ||::: ::::: |||
500 ValMetGlnAspAlaSerGlyGlnHisArgLeuTrpValArgAsnSerG 516
3009 CAGCAACCTGCAGCCTCGAACAATTGACGTTAGTGAAGAAAGACA 3058
      ||::: ||||| ||||| ::::: |||
516 LysSerIleProAlaSerAlaAsnThrLeuLeuValGln..... 529
3059 ACMAACCGCTGTCCGAAAACCTTATTTACCTTCGCAAAACGAA..... 3102
      ||||| ::::: ||||| |||||
530 ...ThrProLeuGlySerAlaIleThrPheThrLeuAlaAsnLysAspL 545
3103 CACGTTCGATCGCGCGCGGTTCACCACTCATCCGCAAAAGACGGCGA 3152
      ||||| ||::: ||||| |||||
546 LysValAspIleGlyThrTyArgTyArgLeu..... 556
3153 GTTCGCGCTGATATTCGGGTCAAAAGACAGAGCTTCGCAACAATCG 3202
556 ..... 556
3203 GCAGAGCAGAACCCAAACAGCGCGGAAAGACACAGCGCAAGCCTT 3252
      ||| ::::: ||||| |||||
557 .....AlaAlaAsnGlyAsnGlyGlnTrp... 564
3253 GACGGCGTGTATTCGCGCGCGCGCGATCCGTCGAAAAGACAGAAAGCCT 3302
      ::::: ||| ||::: |||||
565 ...SerLeuValGlyAla.....Lys 570
3303 TGCAGAACCGCGCGCGGAGGAGGCGGAAATGTCGCATTTATGACAG 3352
      ||| |||||
570 snAlaProProAla..... 574
3353 CGGAGGAAGAGAAAAACGGGTGACGGGATAAAGACACCGCTTCGCG 3402
574 ..... 574
3403 AAACAGCGGAGGAGAAACCGCGCGCTACACCGCTTCGCGCGCGC 3452
      ::::: ||| ||||| |||||
575 .....ProLysProAla...ProLysProLys...ProLys 584
3453 CCGCGCGCGCGCGGATTTGCCGCACTGCAACCCCAACCGCGAGCGCC 3502
      ||||| ::::: ||||| |||||
585 ProProGlnProProGlnProGlnProGlnAla...ProAlaProGlnPro 601
3503 AACCGCAGCGGACCTGTATGACGCTTATGCCAATAGCGGTTTGAGTAA 3552
      ||||| ||||| |||||
601 roAlaGlyArgGlnLeu...SerIleAlaIleAsnAlaIleValAsnThr 616
3553 TTTTCGCGCAGCTCAACAGCGTT.....TTGCGCGTACAGGAGAAAT 3596
```

```
617 GlyIleValGlyLeuAlaSerThrLeuTrpTyAlaGluSerAsnAlaLe 633
      ::::: ||| ||::: ::::: |||
3597 AGACCGGCTATTTCGCGAAGACCGCGCAAC.....GCCGTTT 3634
      ||::: ||||| ||||| |||||
633 uSerLysArgLeuGlyGlyLeuArgLeuAsnProAspAlaGlyAlaTr 650
3635 GGACAAACCGGATCCGGGACACCAACACACTACCGTTGCAAGATTTCGC 3684
      ||| ||||| ::::: |||
650 rPheArgGlyPheAlaGlnArgGlnGlnLeuAspAsnArgAlaGlyArg 666
3685 GCCTACCGCCCAACAAACCGACCTGCGCCCAATCGTATCGCAAAAACT 3734
      ::::: ||| ||::: |||||
667 ArgPheAspGlnLysValAlaGlyPheGlyLeuGlyAlaAspHisAlaVa 683
3735 C.....GGCAGCGGCGCGCTCGGCACTTCCTGTTTCCCAACAC 3772
      ::::: ||| ||||| |||||
683 lAlaValAlaGlyIleArgTrpHisLeuGlyGlyLeuAlaGlyTrpTh 700
3773 GGACCGAAACACCTTCGACGAGCGCATCGCAACTCGCAGCGGCTTCGC 3822
      ||| ::::: ||| ||||| |||||
700 rGlyLysAspArgGlyPheThrGlyAspGlyGlyGlyHisThrAspSer 716
3823 CACGCGCGCGTTTTCGGGCAATACGCATCGACAG...TTCATACATCG 3869
      ||| ::::: ||| ||||| |||||
717 HisValGlyGlyTyArgAlaThrTyIleAlaAspSerGlyPheTyLeuAs 733
3870 CATCAGCGCGCGCGGCTTTAGCAGCGGCACTTC.....T 3907
      ::::: ||||| |||||
733 pAlaThrLeuArgAlaSerArgLeuGlnAsnAspPheLysValAlaGly 750
3908 CAGACGCG.....ATCGAGGCAAAATCCCGCGCGCTGTGATTCAC 3951
      ||||| ||||| |||||
750 erAspGlyTyArgAlaValLysGlyLysTyArgThr.....His 762
3952 GGCATTTCAGCAGCATTCGCGCGCT.....TTCGCGCG 3986
      ||||| ||||| |||||
763 GlyValGlyLysLeuGlyAlaGlyArgArgPheThrHisAlaAsp 779
3987 ATTTCGCGTACGACGCGATCGCGCAACCGCGCTTTCGCAAAAGAG 4036
      ::::: ||||| |||||
779 TyrPheLeuGlnProGlnAlaGlyLeuAlaValPheArgAlaGlyGly 796
4037 CGGATTACCGCTACGAAAC...GTCAATATCGCAACCGCGCTTCGCA 4083
      ::::: ||||| ||||| |||||
796 LysAlaTyArgAlaAlaAsnGlyLeuArgValArgAspGlyGlySer 812
4084 TTCACCGCTACCGCGCGCATTAAGCAGATTTATTCATTCAAACGCGC 4133
      ||| ||||| ::::: |||
813 SerValLeuGlyArgLeuGlyLeuGlyValGlyLysArgIleGlyLeuAl 829
4134 GCMAACATTTTCATCAGCGCTTATTTGAGCTGCTTACGATGCGC 4183
      ||| ::::: ||||| |||||
829 aglyGlyArgGlnValGlnProTyTrpLysAlaSerValLeuGlnGln 846
4184 CTTCGGGCAAAAGTCCGACACGCGCATATACCGCTATTTGGCTCAGAT 4233
      ::::: ||| ||||| |||||
846 heAspGlyAlaGlyThrValHisThrAsnGlyLysAlaArgThrGln 862
4234 TTCGGCAAAACCGGACGTGCGGAAATGGCGGTAAACCGCAATCAAG 4283
      ::::: ||||| ||||| |||||
863 LeuArgGlyThrArg...AlaGlyLeuGly..... 871
4284 TTTTCAGCTGCTCCACGCTGCGCGCGCAAAAGCGCGCAACTGGAAG 4333
      ||||| ::::: ||||| |||||
872 .....LeuGlyMetAlaIleAlaLeuGlyArgGlyHisSerLeuTy 886
4334 CG.....CAACACAGCGCGGCGCATCAATTAGCTACCGCTGCG 4371
      ||| ::::: ||| ||||| |||||
886 lAserTyGlyTySerLysGlyProLysLeuAlaMetProTrp 900
```

seq_name: SwissProt_40:YFAL_ECOLI

CC entitled requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 DR EMBL: A6000313; AAC75293.1; -
 DR EMBL: D90855; BAA16052.1; ALT_INIT.
 DR EMBL: D90854; BAA16050.1; ALT_INIT.
 DR EMBL: K02672; -: NOT_ANNOTATED_CDS.
 DR EMBL: U30459; AAA74054.1; -:
 DR EMBL: Y00544; -: NOT_ANNOTATED_CDS.
 DR Ecogene; EG12850; yfal.
 KW Hypothetical protein; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 1250 HYPOTHETICAL PROTEIN YFAL.
 FT DOMAIN 919 948 15 X 2 AA TANDEM REPEATS OF [DT]-P.
 FT CONFLICT 28 30 AAV -> RGRS (IN REF. 3).
 FT CONFLICT 40 40 K -> Q (IN REF. 3).
 FT CONFLICT 65 66 LV -> PG (IN REF. 3).
 FT CONFLICT 431 431 S -> Q (IN REF. 3).
 FT CONFLICT 434 434 AG -> SA (IN REF. 3).
 FT CONFLICT 478 478 A -> R (IN REF. 3).
 FT CONFLICT 773 773 E -> S (IN REF. 3).
 FT CONFLICT 853 853 V -> M (IN REF. 3).
 FT CONFLICT 923 924 PP -> AT (IN REF. 3).
 FT CONFLICT 948 994 PAYOPYLNAYGVGYLNLRAQAQAFMERRRRHAGGDDQTIN
 FT LRVYGG -> LTRSRC (IN REF. 4).
 FT SEQUENCE 1250 AA; 131152 MW; 1/P9805E299FC95 CRC64;

alignment_scores: quality: 282.00 length: 1445
 ratio: 0.429 gaps: 72
 Percent Similarity: 45.467 Percent Identity: 20.208

alignment block:
 US-09-303-518D-649 x YFAL_ECOLI ..

Align seg 1/1 to: YFAL_ECOLI from: 1 to: 1250

```

637  ACCCGGAAAGTTCATATCATATTGACCGCTATCT..... 675
    ::::|||||:  |||:::~::~|||
46  SerArgInSerLeuSerGly11ethrInSprSer1leAlaSpel 62
676  ....TGGCTGCTGGTGCGCAATACCTTGGCAAAAGATGAGGAGG 721
    |||||  :::::  |||||
62  yGIntrPLeuValPheSerAsp..MetThrAsnAlaSerGly1A 78
722  GCACATCACTTAGTACTGATAAAAAAATTAAACATAGCCCATGCTTT 771
    :::::  |||
78  laval.....Phe 80
772  TTACCAACGAGAGGCTCATTT.....GGGACAGTGGCTC 806
    |||:|||||  |||
81  LeuGlnGlnGlyAlaGluPheSerLeuLeuProGluSlnGlnThrClyPhe 97
807  ACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGTTAATTAATGAGG 856
    :::::  |||
97  tThrLeuPheAlaAsnAsnThr.....ValThrGlyg 108
857  TATTGCAAAAGGAGCAACCCCTATATTAGAAAGAATGAGCTTCAGCTG 906
    :::::|||||:  :::::|||||
108  lntYrAsnAsnGly1yAla1lePheAlaYslnSlnSerThrLeuAsn 124
907  GTTCGTAAGAATGGTCTATGATGAATCTTGGCGAGATACCATTC 956
    ::::  |||  |||||:|||||:
125  LeuThr.....AspVal1lePheSerGlyAsnValAlaGlu 136
957  ACTATTTAGACACACAGTCGAATATGGGAATRCCTTTTAAACGACGATA 1006
    ::::  |||
136  yGly1yTr.....GlyGlyAla1leYrSerSerGlyTr 147
1007  ATATATGACACAGAAATCATCATGCCAAACATGACAC..... 1044
    :|||  |||||  :::::  :::
147  hTrAsnAspTrGlyAlaValAlaPheLeuAlaTrVal1ThrAsnAlMetPheArg 163
  
```

```
1045 AATTCCTGCTATAGATTAAACAGCAACCGTTCAATTGTTAAT.. 1092
|||||
164 AsnAspIleAlaAspIleAspGlySerGlyGlyAlaIleTyrThrIleAsnAs 180
|||||
1093 ...GTTTCTTATCCGAGACAGCA...AGACACCTGTTTTCATG 1132
|||||
180 nasPValTyrLeuSerAspValIlePheAspAsnAsnGlnAlaTyrThrS 197
|||||
1133 CTCGACGTGTGTCACAGCTTATCCAGCCAGACCTGATTAATGAGAAAT 1182
|||||
197 erThSerTyrSerAspGlyAspGlyGlyAlaIleAspValThrAspAsn 213
|||||
1183 ATTTCCTTATTCAGCAAGCAAGCAATGATTAAT... 1221
|||||
214 AsnSerAspSerTyrShsProSerGlyTyrThrIleValAsnAsnThrAl 230
|||||
1222 ...ACCAGCAATCATCAATCAAGTCTGAGGATTAATTTCCAAAGAG 1267
|||||
230 aPherThrAsnAsnThrAlaGlnIleTyrGlyGlyAlaIleTyrThrAsnS 247
|||||
1268 ATTTTACG...GTCCTGCTGAAATACGAACAT 1299
|||||
247 erValThrAlaProTyrLeuIleAspIleSerValAspAspSerTyrSer 263
|||||
1300 TGCGAAGCGCGGCGCTTCATATCAGTAAGACAGTACCGTTACTTGGA 1349
|||||
264 ...GlnAsnGlyGlyValLeuValAspGluAsnAsnSerAlaIleTyr 279
|||||
1350 AGTAACGGGTGGCAACAGCGCGCTTCCAATTCGCAAGCGAC 1399
|||||
279 rGlyAspGlyPro...SerSerAlaIleGlyGlyPheM 291
|||||
1400 TGCACGTTTCAAGCCAAAGGAGAAACCAAGCTCGATCAGCTGGCGGAC 1449
|||||
291 etTyrLeu...GlyLeuSerGlyValThrPheAspIleAlaAsp 304
|||||
1450 GGTACACTCATTTTGGATCAAGCGGAGATTAAGCAAAACAAAGC 1499
|||||
305 GlnTyrThrLeuValIleGlyAsnThrGlnAsnAspGlyAlaValAspSe 321
|||||
1500 CTTAGTGAATTCGCTTGTC...AGCGCAGGCGTACGCTGCAC 1543
|||||
321 rIleAlaGlyThrGlyLeuIleThrTyrThrGlySerGlyAspLeuVal 338
|||||
1544 TGAATGCCGATTAAT... 1557
|||||
338 euAsnAlaAspAsnAsnAspPheThrGlyLeuMetGlnIleGluAsnGly 354
|||||
1557 ... 1557
|||||
355 GluValThrLeuGlyArgSerAsnSerLeuMetAsnValGlyAspThrHn 371
|||||
1558 ...CAGTTCAACCCGCAAACTCTATTTGCGTTCGCGCGGACGTT 1603
|||||
371 scYstIlnsPaspProGlnAspCysTyr...GlyLeuThrIleGlySerI 387
|||||
1604 TGGAT...TTAAACGGGCAATCCGCTTTCG... 1629
|||||
387 leAspGlnTyrGlnAsnGlnAlaGluLeuAsnValGlySerThrGlnGln 403
|||||
1630 ...TTCCACGCTATTCAAATACGATGAGGCGGATGATTCGAA 1673
|||||
404 ThrPheValHisAlaLeuThrGlyPheGlnAsnGlyThrLeu... 417
|||||
1674 CCACATCAAGCAAGAAATCCACCGTTACAT... 1707
|||||
418 ...AsnIleAspAlaGlyGlyAsnValThrValAsnGlnGlySerPheA 433
|||||
1708 ...ACAGCAATAAAGATATGCTACCAACCGCAAT 1740
|||||
433 laGlyIleIleGlnGlyAlaGlyGlnLeuThrIleAlaGlnAsnGly... 448
```

```
1741 AACACAGCTTGATAGCAAAAGAAATGCTCTACACGCGTGGTTGG 1790
|||||
449 SerTyrValLeuAlaGlyAlaGlnSerMetAlaLeuThrGly... 462
|||||
1791 CGAGAAATATACAGCAAAAGAGCGGCTGCACCTGTTTACACAG 1840
|||||
463 ...AspIleValAlaAspAspGlyAlaValIleSerLeuGlnGlyA 477
|||||
1841 CCGCCGCA...GAAGCGCGACCGCTGCTT 1869
|||||
477 sPAlaAlaAspLeuThrAlaLeuGlnAspAspProGlnSerThrIleValLeu 493
|||||
1870 TCCGCGCGAACA...AATTAAACGCAACATCAGCAAAACAGGCA 1916
|||||
494 AsnGlyGlyValIleAspLeuSerAspPheSerThrThrGlnSerGlyTh 510
|||||
1917 ACTGTTTTCAGCGCGACACCAACACCGCAGCTAC... 1953
|||||
510 rSerTyrAsnAspGlyLeuGlnValSerGlySerSerGlyThrValIleG 527
|||||
1954 ...AATCATTTAAACAGCAT 1971
|||||
527 lYSerGlnAspValAlaAspLeuAlaGlyGlyAspAsnLeuHisIleGly 543
|||||
1972 TGCTGCAAAAGAGCGCAT...CTCG 1997
|||||
544 GlyAspGlyLysAspGlyValTyrValAlaValAlaAspAlaSerAspGly 560
|||||
1998 CGGGAATTCGCTGCGCAACGACTGATCAACCGCATTTTAAAGCG 2047
|||||
560 nAlaSerLeuAlaAsnAsnSerTyrLeuGlyThr... 573
|||||
2048 AAACCTTCAATTTAAAGCGGAGCGGCGTGTTCGCAATGTTGCC 2097
|||||
574 ...GlnIleAlaSerGlyThrLeuMetValSerAspAsn... 585
|||||
2098 AATGTAAGCGCATGTCATTTGAGCAATCAGCGCGAGAGTTTGG 2147
|||||
586 SerGlnLeuGlyAspThrHisTyrAsnArg...GlnValIlePheThr 600
|||||
2148 TGTGCGACCGCATCAAG...CACACATCTGTACAC 2182
|||||
600 rAspLysGlnGlnGlnSerValMetGlnIleThrSerAspValAspThr 617
|||||
2183 GTTCGACTGACGGGTGTGACAAATGTGTGCAAAACATTCACGAC 2232
|||||
617 rGSerAspAlaIleGlyHisGlyArgAspIleGln...MetArgAla 631
|||||
2233 GATTAAGTATGCTTCATTCAGCTAAGACGACATCAGCGGCAATGTGCA 2282
|||||
632 AspGlyGluValAlaValAspAlaGlyAlaAspThrGlnThrGlyAlaLe 648
|||||
2283 TCTTCCGATAC...GTCATTTAATC 2308
|||||
648 uMetAlaAspSerSerSerGlyGlnHisGlnAspGlnGlySerThrLeuThr 665
|||||
2309 TCACAGGCTTGCACACTCAACGCGCAATCTTAGTCAATGGCGATTA 2358
|||||
665 yThrGlyAlaGlyThrLeu...GluLeuThrAlaSerGlyThrThr 679
|||||
2359 CGTTATACAGTCAAGCAACAGCCACCAACGCAACCTTAACTGCT 2408
|||||
680 GlnSerAlaVal... 683
|||||
2409 GGGCAATGCCACAGCAACATTTAATCAGCCACATTAACGCGACACAT 2458
|||||
684 ...ArgValGlnGlyThrLeuLysGlyAspValA 695
|||||
2459 CGGCTTCGGCAATGCTTCAATTAATCTAAGCAGCAGCC... 2499
|||||
695 la...AspIleuProTyrAlaSerSerLeu 704
|||||
2500 ...GTACAAACGCGCAGTCTGACGCTTCCGGACGCTAAGCAACGCT 2546
```

705 TrpValGlyAspGlyAlaThrPheValThrGlyAlaAspGlnAspIleG1 721
2547 AAGCATTTCCGCACTC...AAGGTATGTCCTCCATCCGATAGGACAG 2593
721 nSerIleAspAlaIleSerSerGlyThrIleAspIleSerAspGlyThrV 738
2594 TATTCATTTTGAAGACGCCCTTACGGACAAATTCAGCGCGGACAG 2643
738 AlLeu.....ArgLeuThrGlyGln..... 744
2644 GATACGGCATTTACACTTAAGACAC.....GAATGACGCT 2681
745 AspThrSerValAlaLeuAsnAlaSerLeuPheAsnGlyAspGlyThrIle 761
2682 GCGCGACGACGCGAA.....TTAGCAATTTAAAC..... 2712
761 ValAsnAlaThrAspGlyValThrIleuThrGlyGlnLeuAsnThrAsnL 778
2713CTTGACACGCCACATTCATCACTCAAT 2739
778 euGluThrAspSerLeuThrTyLeuSerAsnValThrValAsnGlyAsn 794
2740 TCCGCTATCCGCACGATCGCGACGGGCGCAACCGCATGCGACAGA 2789
795 LeuThrAsnThrSerGlyAlaValSerLeuGlnAsnGlyValAlaGlyAs 811
2790 TCGCGCGCGCGCGCTTCGCGCGCTTCGCGCTCCATATTCGCTTA 2839
811 P..... 811
2840 CACCGCACTTCGTAGATCCCTTTCACACGCTAGCGTAAACGCG 2889
812ThrLeuThrValAsnGly 817
2890 AATTCAGGTCAGGAAATTCGCTTATGTCGAACTCTTCGCG... 2937
818 AspTYrThrGlyGlyThrLeuLeuAspSerGlyLeuAsnGlyAs 834
2938 ...TACCGCAGCAAAATTCAGCTGCGGAAAGTTCCGAAAGCACTT 2983
834 pAspSerValSerAspGlnLeuValMetAsnGlyAsnThrAlaGlyAsn 851
2984 ACACCTTGGCGGTCAACAAT..ACCGGC...AAGCAACCTGCAAGCTC 3027
851 hThrValValValAsnSerIleThrGlyIleGlyGlnProThrSerThr 867
3028 GAACAATTGACGCTAGTGGAAGAAAGACAACAACCGCTGCGGAAA 3077
868 GlyIleValValValAspPheAlaIleAspProThrGlnPheGlnAsnAs 884
3078 CCTTAATTTACCCGCAAAAGCAA...CACGTGATGCGCGCGCTGCG 3124
884 nAlaGlnPheSerLeuAlaGlySerGlyTYrValAsnMetGlyAlaIleTYr 901
3125 GTTACCAATCATCCGCAAAAGCGGAGTTCCGCTGATTAATCCGCTC 3174
901 sPTyTYrThrLeuValGlnAspAsnAsnAspTYrLeu..... 913
3175 AAGACAAAGAGCTTTCGACAAACTCGGACAGGACGAACCAAAAAACA 3224
913 913
3225 GCGGAAAAAGACAAGCGCAACGCTTGAACGGCTGATTCGCGCGGCG 3274
913 913
3275 GCGATGCCGTCGAAAGACAGAAAGCGTTCGCAACCGCGCGCGACAGA 3324
914ArgSerGlnGluValThrProProser..... 922
3325 GCGCGGAAAAATGCGCATTTATGCAAGCGGACGAGAAAGAAAAACGGGT 3374
922 922
3375 GCAGCGGATAAAGACACACCGCTTGGCGAAACAGCGCGGAGGAAACCC 3424
923 ..ProProAspProAsp..... 927
3425 GCGCGGCTACACCGCTTCCCGCGCGCGCGCGCGCGCGGATTTG 3474
928 ..ProThrProAsp..... 931
3475 CCGCAATGTCACCCCAACCGGACCGCCCAACCGGACGACCTGATAG 3524
932 ProAspProThrProAspProAspProThrProAspProGluProThrPr 948
3525 CCGTATGCCAATAGCGGTTTGAGTAATTTCCGCAACCGCGCACACG 3574
948 oAlaTYrGlnProValLeuAsnAlaValAlaGlyGlyTYrLeuAsnAsnL 965
3575 TTTTCGCGCTACAGACGCAATTCAGCCGCTATTTGCCGAACCGCGCC 3624
965 euArgAlaAla.....AsnGlnAlaPheMetMetGluArgTYr 977
3625 AACGCGCTTGGACACGCGC.....ATCCGGGA 3653
978 AspAlaIleGlyGlyAspGlyGlnThrLeuAsnLeuArgValIleGlyG1 994
3654 CACCAACACATACCGTTCCGAAGATTCGCGCTACCGCAACAAACCG 3703
994 yAspTYrHisTYrThrAlaAlaGlyGlnLeuAla.....GlnHisGlu 1009
3704 ACCGTGCGCAATCGGTATGCGAAGAAACCTCGGACGCGCGCG... 3747
1009 sPTThrSerThrValGlnLeuSerGlyAspLeuPheSerGlyArgTYrGly 1025
3748GTGCGCATCCGTG...TTTGCGCAACCG 3773
1026 ThrAspGlyGluTYrMetLeuGlyIleValGlyGlyTYrSerAspAsnG1 1042
3774 GACGCAAAACACCTTCGACGACGCGCATCGGCACTCGGACGCGTTGCC 3823
1042 nGlyAspSerArgSerAsnMetThrGlyThrArgAlaAspAsnGlnAsn 1059
3824 ACGGCGCGCTTTCGCG.....CAATCGGATCGAC 3855
1059 IsGlyTYrAlaValGlyLeuThrSerSerTYrPheGlnHisGlyAsnGln 1075
3856 AGTTCATCATCGCGCATGACGCGCGCGCGGCTTTAGC.....ACCG 3899
1076 LysGlnGlyAlaTYrPheAspSerTYrPheGlnTYrAlaTYrPheSerAs 1092
3900 CAGCTTCGACGCGCATCGGAGGCAAAATCCCGCGCGCTGCTGAT 3949
1092 nAspValSerGlnGlnGluAspGly.....ThrAspHis 1104
3950 AC.....GGCATTCAGGACGATACCGCGCGCGCTTC... 3981
1104 yHisSerSerGlyIleAlaSerLeuGlnAlaGlyTYrGlnTYrPhe 1120
3982GGCGATTCGCGCATCGAACCGCAATCGGCGCAACGCGTATTT 4025
1121 ProGlyArgGlyValValIleGlnProGlnAlaGlnValIleTYrGlnG1 1137
4026 CGTCAAAAGCGGATTCGCTACGAAAC..... 4056
1137 yValGlnGlnAspAspPheThrAlaAlaAsnAlaGlyAlaTYrSerGlnS 1154
4057GTCAATATGCGCAACCGCGCGCTTCATTCACCGTACCGC 4098
1154 erGlnGlyAspAspIleGlnThr.....Arg 1162
4099 GCGGCATTTAAGGAGAGATTTATTCATTCAAACGCGCGCAACATTTCCAT 4148
1163 LeuGlyLeuHisSerGluTYr.....ArgThrAlaValHis.....Va 1175

```

4149 CAGCCTTATTGAGCCTGTC...TATACGATGCCGCTCGGCAAG 4195
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1175 lIleProThIleuAspLeuAsnTyTrHIsAspProHISerThIglut 1192
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4196 TCCGAACACGCGTCATACCCTGTTGGCTCAGATTTCGCAAAAC 4245
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1192 IeGIuGluAspGIserThr.....IleSerAspAlaValAlaValSglu 1206
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4246 CGC...AGTCCGGATGGCGCTAAACGCCAATCAAGGTTTCACGCT 4292
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1207 ArgGIuGIuIleIleValGIuValThrGIuAsnIleSerGln...ArgVa 1222
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4293 GTCCCTCCAGCGTCCGCCGCC.....AAAGCCCGCAACTGGAAAGCC 4336
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1222 lIserIleuArgGIserValAlaIleTrpGlnIleValGIserAspPheAlaG 1239
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4337 AACACAGCGCGGCAATCAATAGGCTACCGCTGG 4371
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1239 lThrAlaGIuPheIleuSerMetThrValIleStry 1250
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq_name: swissprot_40:AIDA_ECOLI

seq_documentation_block:

ID AIDA_ECOLI STRAND: PR: 1286 AA.

AC 003155;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last annotation update)

DE Adhesin aidA-I precursor.

GN AIDA-I.

OS Escherichia coli.

OG plasmid p186.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

NCBI_Taxid=562;

OX [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.

RC STRAIN=2787 (O126:H27);

RC MEDLINE=9232658; PubMed=1625582;

RA Benz I., Schmidt M.A.;

RT "AIDA-I, the adhesin involved in diffuse adherence of the

RT diarrhoeagenic Escherichia coli strain 2787 (O126:H27), is

RT synthesized via a precursor molecule.";

RL Mol. Microbiol. 6:1539-1546(1992)

- FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE

ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI

TO EPITHELIAL CELLS.

- SUBCELLULAR LOCATION: Outer membrane.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

alignment_scores:

Quality: 275.00

Ratio: 0.434

Percent Similarity: 43.416

Length: 1458

Gaps: 65

Percent Identity: 18.999

alignment_block:

US-09-303-518d-649 x AIDA_ECOLI ...

Align seg 1/1 to: AIDA_ECOLI from: 1 to: 1286

```

688 GCATATACCTTTGCAAAATGATCAGT.....GTGGCAC 725
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
45 GlySnaIaPheAlaValAlaAsnIleSerGIyThrValSerSerGIyGlyTh 61
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
726 AGTCACCTAGTAGTGAAAAATTAAACATAGCCCATATGCT..... 768
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
61 rValSerSerGIyGluThrGlnIleValAlaValSerGIyArgGIySnsrA 78
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
769 .....TTTTACCAACAGAGGC... 786
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
78 SnaIaThrValAsnSerGIyGlyThrGlnIleValAsnSngIyGIyLys 94
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
787 .....TCATTGGCGACAG 800
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 ThrThrAlaThrThrValAsnSerSerGIySerGlnAsnValGIyThrSe 111
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
801 TGGCTCACCAATG.....TTATCTATGATGCCCAAAACCAAGTGTG 844
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
111 rGIyAlaThrIleSerThrIleValAsnSerGIyGIyIleGlnArgValS 128
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
845 TAATTATATGGGTATTGCAACGGGCAACCCCTATATGAAAAAGCAAT 894
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
128 rSerGIyGIyValAlaIleSerAlaThrAsnLeuSerGIyGIyAlaGlnAsn 144
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
895 GCCTTCACAGCTGTTCGTAAGATGTGTTCTATGATGAATCTTGTGTG 944
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 lIeTyAsnLeuGIyHISAlaSer.....AsnThrValIlePheSerGI 159
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
945 AGATACCATTCAGTATTC..... 963
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
159 yGIyAsnGlnThrIlePheSerGIyGIyIleThrAspSerThrAsnIleS 176
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
964 .....TACGACCCAGCTCAAAATGGCAATFACCTTTTACGAC 1002
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
176 rSerGIyGIyGlnGlnArgValSerSerGIyGIyValAlaIleSerSnh 192
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1003 GATTAATATGGCAAGAAAAATCAATGCCCAACATGACACATCTCT 1052
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
193 ThrIleAsnSerSerGIyAlaGlnAsnIleLeuSerGIyGlnGlyAlaI 209
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1053 GCCTAATAGATTAAACACAGAACCGTCAATGTTTAACTTTCTTAT 1102
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
209 rSerThrHISIleSerSerGIyGIyAsnGlnIlyr.....IleSerAlaG 224
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1103 CCGAGACACAGACAGAACTGTTATCATCGTCGACAGTGTCAACAGT 1152
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
224 lYAlaAsnAlaThrGlnThrIleValAsnSerGIyGIy..... 236
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1153 TATCGACCCAGCTGATAATGCA..... 1176
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
237 ...PheGlnArgValAsnSerGIyAlaValAlaThrGIyThrValIleuSe 252
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1177 .....GAAATATTTCCTTTTATGACGAGAGAAAAAGCGGATGTA 1216
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
252 rGIyGIyThrGlnAsnValSer.....SerGIyIleSerAlaI 265
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1217 TACTTACGACCAACATCAATCAAGTGTGAGAGATATATTTCCAGAGA 1266
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
265 lSerThrSerValIlyrAsnSerGIyValGlnThrVal...PheAlaGIy 280
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1267 .....GATTTTACGCTGTGCTCGCTGAAAAATRACGAAACTGTGCA 1304
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
281 AlaThrValThrAspThrThrValAsnSerGIyGIyAsnIleSnlIeSe 297
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1305 AGCGCGCGGTTCATATCAATGAGACACTTACCTTACTTGGAAAGTAA 1354
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
297 rSerGIyGIyIle...ValSerGlu.....ThrThrValAsnValS 310
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1355 ACGGCGTGCACAAAC.....GACCGCTGTCCAAATCGGC 1389
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

310  ....: |||
    310  erglythrlnasniletyrserglytyrlealeuSerAlaasnile 326
1390  AAGGACGCGTCACGCTTACAGCAAGGAA..... 1422
327  LysglyserlnilevalasnsergllythrAlaIleasnthrleuVa 343
1423  .....AACCAAGGCTGCATCAGCGGGG 1447
343  lseraspglytyrlyrlnhstleargasnlyglylleAlaSerGly. 359
1448  ACGGTACACTATTGATCAGACGAGCATTAAGCAAAAAACAA 1497
360  .....ThrIleValasnlnserglytyrValasnIleSerSerGly 373
1498  GCCTTAGTGAATCGCTTGTCACGCGGACGAGGTACGCTGCACTGA 1547
374  GlytyrAlaIleuSerThrIleIleasnserglytyrleuArGlylle 390
1548  TCGCGATTAATCAATTCACCCGACAACTTATTCGGCTTCGGGCG 1597
390  uSeraspGlytyrAlaArgGlyThrIleuasn.....AsnSerG 404
1598  GACGTTGATTTAAAGCGGCAATTCGCTTCACCGTATCAAAAT 1647
404  LyrArgIleuasnValSerasnlyglyValSerTyrsnAlaMetIleasn 420
1648  ACCGATGAAGGGCGCATGTCATTCACCAACCAAGCAAGAAATCCAC 1697
421  ThrGlytyrAsnlnlyrIleTyrsaspGlyIleAlaThrAlaIle 437
1698  CGTTACCATTAACAGCAATTAAGATATTCCTCAACCGCAAT..... 1740
437  eValasnThrserGlyPheGlnArgIleasnserglytyrAlaIlePro 454
1741  ..AACACAGCTTGATAGCAAAAAAAGAAATGCC..... 1773
454  alGlnasnserValValThrArgThrValSerSerAlaAlaIlePro 470
1774  .....TACAAAGCT..... 1782
471  PheAspAlaGlnValTyrseryglytyrGlnThrValTyrlleuThrPar 487
1783  .....TGCTTGGCGAGAAAGATACG..... 1803
487  gGlyIleTyrsasnPheleuThrAlaValThrPserMetPheProG 504
1804  .....ACCAAAAGCAAGCGCGGCTCAAC..... 1827
504  LyrThrAlaSerGlyAlaasnValasnleuSerGlyArgleuasnAlaPhe 520
1828  .....CTTGT 1832
521  AlaGlyasnValValGlyThrIleleuasnGlnGlnGlyArgGlnTyVa 537
1833  TTACAGCGCGCGCGCA.....GAGACCGCA 1858
537  lTyrseryAlaThrAlaThrSerThrValGlyAsnasnGlnGlyArg 554
1859  CCCTGCTCTTCGCGCGCAACAATTTAAAGCGCAACATCAGCAAAAC 1908
554  lutyValleuSerGlyIleThr.....AspGlyThrValleuasnSer 569
1909  AAGCGCAACATGTTTTCAGC.....GGACAGCAACACCGCAGCCTA 1952
570  GlytyrleuGlnAlaValSerSerGlytyrLysAlaSerAlaThrValI 586
1953  CAATCATTTAAAGCAACATTTGTCGCAAAAAGAGGCAATTCCTCGCGG 2002
586  easn.....GlyGly.....GlyAlaG 592
2003  AATTCGTGTGGACAAACGATGATCAACCGACATTTAAAGCGGAAAC 2052
    :.....: |||
592  lnpheValTyrsPglytyrGlnValThrGlyThrAsnIleLysasnly 608
2053  TTCCAAATTTAAAGCGGACAGCGGTGTTCCGCAATGTTGCC...AA 2099
609  GlyThrIleArgValAspSerGlyAlaSerAlaIleuasnIleAlaLeu 625
2100  AGTGAAGCGCATTTGGCATTTGACCAATCAGCC..... 2133
625  rserGlytyrAsnleuPheThrThrGlyAlaThrleuProGlnleuThr 642
2134  ....CAAGCAGTTTGTGTGCGACCG...CATCAAGCGCACACATC 2175
642  hrThrMetAlaIleuSerValSerGlnasnIleAlaSerAlaIleVal 658
2176  TGTACACCTTGAGACTGACGCGGTCTGCAAAATTTGTGCAAAAACCAT 2225
659  leuGlnasnlyglytyrleuArGValThrSer.....GlytyrThrAl 673
2226  TACGAGCATTAAGTATGCTTCATTCATTAACACGACATTCAGCGCA 2275
673  thrAsp.....ThrThrValasnSerAlaGlyA 683
2276  ATGTGATCTTGCCGATCACGCTCATTTAAATCTCACAGGCGCTGCACA 2325
683  rGlnArGlyIleAspAspGlytyrThrIleasn.....GlyThrThrThr 697
2326  CTCACAGGC..... 2334
698  IleasnIleAspGlyIleValAlaGlyThrAsnIleGlnAsnAspGlyAs 714
2335  .....AATCTTAGTGAATGGCGATACGCTTATACAGTCAAGCC 2374
714  nPheIleleuasnleuAlaGlnAsnTyrsPheGlnThrGlnleuSerG 731
2375  ACAAGGCC.....ACCAAAAGCGC 2394
731  LyseryValleuValLysaspasnThrGlyIleMetThrTyAlaGly 747
2395  AACCTTAGCCTC.....GTGCGCAATGCCCAACCAACAT 2429
748  ThrleuThrGlnAlaGlnGlyValasnValLysasnlyglyIleIlePhe 764
2430  TTAATCAAGCCACATTTAAAGCGCAACACATCGGCTGCGGCAATGCTCAT 2479
764  eAspSerAlaValValasnAlaAspMetAlaValasnGlnAsnAlaTyrl 781
2480  TTAATCTAAGCGACCGCGCTACAAAAGCGAGCTGACGCTTTCGCGC 2529
781  leasnIleSeraspGlnAlaThrIleasnlyser.....ValasnAsn 795
2530  AAGCGTAAAGCAACGTAAGCCATTCGCGCACGTCGACGTAATGTCGCC 2579
796  AsnlyserIleValIleasnasnSerIleIleIleasnlyAsnIleThr.. 811
812  ....AsnAspAlaAspIleuSerPheGlyThrAlaLys..... 822
2630  TCAGCGCGCGCAAGATACGCGATTACACTTAAAGAAGCAAGCAATGACG 2679
822  ..... 822
2680  CTGCGCTCAGCAGCAAGATTAAGCAATTTAAACCTTGACACGCGCCAT 2729
823  leuIleuSerAlaThrValasnlyserleu...ValasnIleLysasnIle 838
2730  TACACTCAATTTCCGCTATTCGACAGATGCGGAGGCGCAAAACCGCGCA 2779
838  eIleleuasn..... 841
2780  GTGCGACAGATGGCGCGCGCGCGCTTCGCGCGCTTCGCGCGCTTCCTA 2829
841  ..... 841

```

2830 TTATCCGTTACCCGCAACTGCGTAGAATCCCGTTTCAACAGCTGAC 2879
 842ProthrLys...GluSerAlaGlyAsnThrLeuTh 852
 2880 GGTAAACGCGCAAAATGTAACGCGTACAGGCAATCCGCTTATATGCGGAC 2929
 852 rValSer.....AsnTyrThrGlyThrProGlySerValIleSerL 866
 2930 TCCTCGGCTAC.....CGCAGCGACAATGTAACCTG 2961
 866 euGlyGlyValLeuGluGlyAspAsnSerLeuThrAspArgLeuValVal 882
 2962 GCGGAATCTCCGAAGGCACTTACACCTTGGCG...GTCAACATACCG 3008
 883 LysGlyAsnThrSerGlyInSerAspIleValTyrValAsnGluAspG 899
 3009 CACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGAAAGACA 3058
 899 ySerGlyGlyGlnThrArgAspGlyIleAsnIleIleSerValGluGlyA 916
 3059 ACAACCGCTGTCCGAAAACCTTAATTTACACCTGCAAAAGCAACAGCTC 3108
 916 sn.....SerAspAlaGluPheSerLeuLysAsn...ArgVal 927
 3109 GATGCGGCGGCGTGGCGTTACCACTCATCCGCAAGAGCGGAGATTCCG 3158
 928 ValAlaGlyAlaTyrAspTyrThrLeu...GlnYsglyAsnGluSerG 943
 3159 CCTGCATATCCGTCGTAAGAAACAAGAGCTTTCGCAACACTGCGCAAG 3208
 943 yThrAspAsn.....LysGlyTyrPtyrLeuThrSerHisLeuProThr 958
 3209 CAGAAGCCCAAAAACAGCGGCAAAAGCAACGCGCAAGCCTTGACCGC 3258
 958 eAspThrArgGlnTyrArgProGluAsnGlySerTyrAlaThrAsnMet 974
 3259 CTGATTGCGGCGCGGCGGATCCGTCGAAAAGACGAAGAGCTTGCGCA 3308
 975 AlaLeuAla..... 977
 3309 ACCGCGCGCGAGCGAGCGGGAATGTGCGCATTTATGCAAGCGGAGG 3358
 978AsnSerLeuPheLeuMetAspLeuA 986
 3359 AAGAAAAAAGCGGTGCGAGCGGATTAAGACACCGCCTTGCGCAACAG 3408
 986 snGluArgLysGlnPheArgAlaMetSerAsp..... 996
 3409 CGCGAAGCGGAAACCGCGGCTTACACCGGCTTCCCCCGCGCGCGCG 3458
 997AsnThrGlnProGluSerAla..... 1003
 3459 GCGCGCGCGGAGTTGCGCAACTGCACACCCCAACGCGACCCCAACGC 3508
 1003 1003
 3509 AGCGGACCTGATCAGCCGTATGCCAATAGCGGTTAGTGAATTTCC 3558
 1003 1003
 3559 GCCACGCTCACACAGCGTTTCCCGCTACAGACGAAATTAGACCGGTA 3608
 1003 1003
 3609 TCGCGAAGACCGCGCAACGCGGTTGG.....ACAAGCGGATCC 3649
 1004SerValTrpMetLysIleThrGlyGlyIle 1014
 3650 GGGACACCAACACTACCGTTGCAAGATTTCGCGCTACCGCCACACAA 3699
 1014 eSerGlyLys..... 1017

3700 ACCGACCTCGCCAAATCGTATGCAAAAAAAGCTCGGAGCGGCGCGT 3749
 1018LeuAsnAspGlyLys... 1022
 3750 CGGCACTCTGTTTTCGACACACCGGCAAAACACTTTCGACAGCGCA 3799
 1023AsnLysThrThrAsnGlnPheIleAsnGlnL 1034
 3800 TCGGCACACTCGGACGCGCTTGGCCACGCGCGCTTTCGGGCAATACGC 3849
 1034 euGlyGlyAspIleTyrLysPheHisAlaGluGlnLeuGlyAsp..... 1048
 3850 ATCGACAGGTTCTACATCGCATACAGCGGCGCGGCGGTTTATGACGCG 3899
 1049PheThrLeuGlyIleMetGlyGlyTyrAlaAsnAlaLysG 1062
 3900 CACCTTTACAGACGCGATCGAGGCAAAATCCGCGCGCGCTGCTGAT 3949
 1062 yLysThrIleAsnTyrThrSerAsnLysAlaAlaArgAsnThrLeuAsp 1078
 3950 ACGCATTCAGGACGATACCGCGCGGCTTTCGGCGGATTCGCAATGCA 3999
 1079GlyTyrSerValGlyValTyrGlyThrTyrPtyrGln 1090
 4000 CCGCACATCGCGCAACGCGCTATTTCTGCAGAAA...CGCGATTACCG 4046
 1091 AsnGlyGluAsnAlaThrGlyLeuPheAlaGluThrTyrMetGlnTyrAs 1107
 4047 CTACGAAACGTCATATATCGCCGCCCGCGCTGATTCACACGCTTAC 4096
 1107 nTrpPheAsnAlaSerValLysGlyAspGlyLeuGluGlnGlyLysTyrA 1124
 4097 GC..... 4098
 1124 snLeuAsnGlyLeuThrAlaSerAlaGlyGlyTyrAsnLeuAsnVal 1140
 4099GCGGCATTAAAGCAGATTAATTCATTCAAACG 4130
 1141 HisThrTyrThrSerProGluGlyIleThrGlyGlnPheTrpLeuInPr 1157
 4131 G.....GCGCACACATTCATCAGCGCTTATTTAGCGCTGCTCT 4171
 1157 OHISLeuGlnAlaValTrpMetGlyValThrProAspThrHis..... 1171
 4172 ATACCGATGCGGCTTCGCGCAAGTCCGAACGCGCTCAATACGCGCTA 4221
 1172GlnGluAspAsnGlyThr 1177
 4222 TTGGCTCAGATTTTCGCAAAACCGCG.....AG 4250
 1178 ValValGlnGlnAlaGlyLysAsnAsnIleGlnThrLysAlaGlyLeuA 1194
 4251 TCGCGAATGCGGCTTAACGCGCAAAATCAAAAGTTTACGCTGCTCC 4300
 1194 gAlaSerTrpLysValLysSer.....ThrLeuAspLysA 1206
 4301 ACCTGCGCGCGCGCAAGCGCGCAACTGGAACGCA..... 4338
 1206 sPthrGlyArgArgPheArgProTyrIleGlnAlaAsnTrpIleHisAsn 1222
 4339 ...CACAGCGGCGCATCAATTA 4359
 1223 ThrHisGluPheGlyValLysMet 1230

seq_name: SwissProt_40:YEJO_ECOLI
 seq_documentation_block:
 ID YEJO_ECOLI STANDARD: PRT: 863 AA.
 AC P33924, P76450: Q47291,
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical outer membrane protein yeJo.

GN YEJO OR B2190.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae;
 OC Escherichia.
 OC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / BHB2600;
 RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
 RA Church G.M.;
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Berra N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP PRESENCE OF AN INSERTION SEQUENCE.
 RA Rudd K.E.;
 RL Unpublished observations (JAN-1994).
 CC -1 SUBCELLULAR LOCATION: Outer membrane (Potential).
 CC -1 SIMILARITY: STRONG, TO BORDETELLA PERTUSSIS.
 CC -1 CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE READING FRAME IS
 CC INTERRUPTED BETWEEN CODONS 21 AND 22 BY A 155K INSERTION ELEMENT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U00008: AAL16385.1: ALT-SEQ.
 DR EMBL: AE000308: AAC75250.1: ALT-SEQ.
 DR Ecogene; Egi2051: yejo.
 KW Hypothetical protein; Outer membrane; Complete proteome.
 SEQUENCE 863 AA; 91202 MW; 55D4600B5BC3D94D CRC64;

alignment_scores:
 Quality: 265.50 Length: 1023
 Ratio: 0.575 Gaps: 49
 Percent Similarity: 45.161 Percent Identity: 20.723

alignment_block:
 US-09-303-518D-649 x YEJO_ECOLI ..

Align seg 1/1 to: YEJO_ECOLI from: 1 to: 863

```

1693 TCCACCGTT.....ACCATACAGC.....ATAAGA 1721
|||||
30 SerthrValGluTyrGluThrValAspGlyValValLeuIleuIle 46
1722 TATGCTACA.....ACGGCATAAACACAGCTTGATAGCAAAA 1762
|||||
46 PileGlnLeuValTyrGlyThrAlaAsnThrLysIleAsnProGly 63
1763 AAGAAATGGCTACACAGTTGGTTGGCGAGAAAGATACAGCAAAACG 1812
|||||
63 LysIleu...GlnHisIleLysGluPheGlyValSerAsnAsnThrGlu 78
1813 AACGGCGGCTCAACCTTTGTTACAGCCGCGAGAGACCGACCCCT 1862
|||||
79 AsnGlyTyrGlnTyrIleGluMetAsnGlyAlaIleGluTyrSer 95
1863 GCTGCTTCGGCGGA.....ACAATTAACGGCAACATCAGCAAA 1906
|||||

```

```

95 IleAsnAspGlyTyrGlnIleValGlnMetGlyAlaIleAsnGln 112
1907 CA.....AACGCAAACTGTTTTCAGGGGAGACCAACACCGAC 1947
|||||
112 hrThrIleuAsnAsnGlyValLeu.....Gln 120
1948 GCCTACAAATCAATTAACAGCATTGTCGCAAAAAGAGGCAATTCCTG 1997
|||||
121 ValTyrGlyAlaAlaAsnAsp.....ThrThrIleIleuSgl 132
1998 CGGGAAATCGTGTGGACACAGACTGATACACCGCATTTAAAGCG 2047
|||||
132 YelTyrGluLeuIleValGluLysAspGly...GlyAlaValPheValAla 148
2048 AAAACTTCCAAATTAAGCGCGAGCGGCGTGTCCCGCAAT..... 2091
|||
148 IeGlu.....LysGlyGlyLeuLeuGluValLysGluGlyPhe 161
2092 .....GTGCCAAAGTGAAGCGCATTTGGCAATCA 2129
|||||
162 AlaPheAlaValAspGlnLysAlaGlyAlaIleLysThrThrThr 178
2130 CGCCCAAGCAGTTTGTGTC..... 2151
|||||
178 GAlaMetGluValPheGlyThrAsnArgLeuGlyLysPheAspIle 195
2152 .....GCA 2154
|||
195 snglyTLeAlaAsnAsnMetLeuLeuGluAsnGlyLysLeuArgVal 211
2155 CGGCATCAAGCGCACCAATCTGACGTCGAGACGAGCGTCTGAC 2204
|||||
212 GluGluAsnAspPheAlaTyrAsnThrValAspSerGlyLeuLe 228
2205 AAATGTGTCAA.....AAACCATTACC.....GACG 2233
|||
228 uGluValMetAspGlyGlyThrValThrGlyValAspLysAlaGly 245
2234 ATAAAGTATGCTTCATTCATTGACTAAGACGACATCAGCGCAATGTCAT 2283
|||||
245 IlyLysLeuIleValSerThrAsnAlaLeuGluValSerGlyPro..... 259
2284 CTGGCCGATCAGCGCTCAATTTAATCTCAGAGCGCTCCACACACG 2333
|||||
260 .....AsnSerLysGlyLysPheSerIleLysAs 269
2334 CAATCTTATGTCAAAT.....GGCGATACAGCTTATACAGTCAGCC 2374
|||||
269 pGlyValSerLysAsnTyrGluLeuAspAspGlySerGlyLeuIleVal 286
2375 ACAAGCGCACCCCAAGGCGACCTTGCTGCGCAATGCCCAAGCA 2424
|||
286 etGluAspThrGln.....AlaIleAspThrIleuAspLysHisAla 300
2425 ACATTTATCAAGCCACATTTAAACGGCAACATCGGCTTGGGCAATGC 2474
|||||
301 ThrMetGlnSerLeuGlyLysAspThrGlyThrLysValGlnAlaAsn 317
2475 TTCAATTAATCAAGCAGCAGCCGCTGACAAAGCGCACTGACGCTT 2524
|||||
317 aValTyrAspLeuGly...ArgSerTyrGlnAsnGlySerIleThrTyr 333
2525 CC.....GGCAACGCTTAAGCAAC 2544
|||
333 eSerLysAlaIleSerGluAsnMetValIleAsnAsnGlyArgAlaAsn 349
2545 GTA.....AGCATTCGCACTCAACAGGTAATGTCCTCCT 2579
|||||
350 ValTyrPalaGlyThrMetValAsnValSerValArgGlyAsnAspGly 366
2580 AGCGGAT.....AAGCAGTATTCATTTTGAAGACCGCGCTTACCG 2623
|||
366 eLeuGluValMetLysProGlnIleAsnTyrAlaProAlaMetLeuVal 383

```

2624 GACAAATC.....ACGGCGGCAAG 2643
|||:::|
383 LylsValValValSerLugLylAlaSerPheArgThrHisGlyAlaVal 399
2644 GATACGGCA.....TTACACTTAAAGACAGGAAATGACCGTGC 2684
|||||:::|
400 AspThrSerLysAlaSpValSerLeuGluAsnSerValTrpThrLeu 416
2685 GTCAGGC.....ACGGAATTAGGCAAT 2707
|||:::|
416 eAlaAspLleThrThrThrAsnGlnAsnThrLeuLeuAsnLleAlaAsnL 433
2708 TAAACCTTGACAAAGCCACCATTACATCAATCCGCTATGCGCCAGAT 2757
|||:::|
433 eulMetSerAspAlaAsnValLleMet..... 442
2758 GCGGCAAGGGCGCAACCGGCACTGGACAGATGCGCGCGCGCTTC 2807
443MetAspGluProValThrArgSe 450
2808 GCGCGGTCGCGCGCTTATCCGTTACACCGCCAACTCGGTAG 2857
|||:::|
450 r.....SerValThr.....AlaSerAlaG 457
2858 AATCCCTTCAACAGCGCTACGCTAACCGCAATTTGAAGGTACGGGA 2907
|||:::|
457 Lysn...PheLleThrLleThrAsn...ThrLeuSerGlyAsnGly 471
2908 ACATTCGCTTTATGTGGAACCTTTCGCTACCGGACGACAAATTTGA 2957
|||:::|
472 AsnThrLysMetArgThrAspMetAlaAsnHisGlnSerAspGlnLeuAs 488
2958 GCTGGCGGAAAGTTCCGAGGCACTTACACCTTGGCGGTACAAATACG 3007
|||:::|
488 nValThrGlyGlnAlaThrGlyAspPheLysLlePheValThrAspThrG 505
3008 GCAACGAAAGCTCGAAGCCTGCAACATTTGACGTAAGTGA 3051
|||:::|
505 LylAsnProAlaLysLysSerLeuThrLeuValThrThrGlyGly 521
3052 AAAGACAACAACCGCTGTCGAAACCTTAAATTCACCTGCAAAACGA 3101
|||:::|
522 GlyAsp.....AlaLalPheThrLeuGlyAsnAl 531
3102 ACAC.....GTGATGCGGCGGTGGCGTTACCACTCATCGCAAG 3145
|||:::|
531 agLylGlyValValAspLleGlyThrTyGlyLysThrLeuLeu..... 545
3146 ACGGGAAGTTCGGCTGCATATCCGCTCAAGAAAGACGTTTCGAC 3195
545 545
3196 AAACCTCGCAAGCGCAAGCAAAAGACGCGGCAAGCAAGCGGCA 3245
|||:::|
546AspAsn..... 547
3246 AAGCCTTACCGGCTGATTTGGCGCGGCGGCGATGCGTGAAGAAAGACAG 3295
547 547
3296 AAAGCGTTGCGAAGCGGCGCGGCGAGCGAGCGGGGAAATGTGGCATTT 3345
547 547
3346 ATGCAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3395
548GlyAsnHisSerTrpSe 553
3396 CTTGGCGAAGACAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3445
|||:::|
553 rLeuAlaGlyLysnArg...AlaGlnLleThrProSerThrThr..... 566
3446 CCGCGCGCGCGCGCGCGCGCGCGGATTTGCCGCAACTGCAACCCCAACG 3495
566 566
3496 CAGCCCAACCGCAGCGGACCTGATCAGCGGTTATGCCAATAGCGGTT 3545
|||:::|
567AspValLleuAsnMetAlaLalAlaLalGlnProLe 577
3546 GAGTGAATTTCCGCAAGCGTCAACAGCGTTTTCGCGTACAGAGCAAT 3595
|||:::|
577 u...ValPheAspAlaGluLeuAspThr.....ValArgGluArgL 590
3596 TAGACCGGCTATTGTCGAGAGACCGCGCAAGCGGCTTTGGACAGAGCGC 3645
|||:::|
590 euGlySerValLysGlyValSerTyAspThrLametrPserSerAla 606
3646 ATCCGGGACACCAACACTTACCGTTCCGCAAGATTTCCGCGCTACGCCCA 3695
|||:::|
607 Ile...AsnThrArgAsnAsnValThrThrAspAlaGlyLysPheG 622
3696 ACAAAC.....GACCTGCGCCAAATCGGTA 3721
|||:::|
622 uGlnThrLeuThrGlyLeuThrLysGlyLleAspSerArg.....P 636
3722 TGCAGAAAACCTCGCAGCGGCGCGCTGCGCATCTGTTTGCACAGAC 3771
|||:::|
636 heserArgLysLysSerThrLleArgGlyLysLlePheGlyLysSer 652
3772 CCGACCGAAACACTTACGACGCGCATCGCAACTGCGCAGCGGCTGC 3821
|||:::|
653 HisSerAspLleGlyPheAspArgLys..... 661
3822 CCACGCGCGGCTTTTGGGCAATACGCAATGCAAGGTTTACATCGCA 3871
|||:::|
662GlyLysGlyAsnLleAspSerTyThrLeuGly 673
3872 TCAGCGCGGCGCGGCTTTACGACGCGGCGGCTTACAGCGC..... 3915
|||:::|
673 lATrAlaGlyLysPheLysGlnAsnGlyAlaTyValAspLysValAl 689
3916ATCGAGGCAAAATCGCGCGCGCT 3941
690 LysValAspArgPheAlaAsnThrLleHisGlyLysMetSerAsnGlyAl 706
3942 GCTGATTAAGC.....ATTCAG 3961
|||:::|
706 arThrAlaPheGlyAspTyAsnSerAsnGlyAlaGlyAlaHisValG 723
3962 CACGATACGCGCGGCTTTCGCGGATTCGCGATGCAACCGCAATCGCG 4011
|||:::|
723 erGlyPheArgTyPheValAspGlyLeuTrpSerValArgProTyLeuAla 739
4012 GCAACGCTATTTCGCAAAAGCGGATTAACCGCTTACGAAAGCTCA 4061
|||:::|
740 PheThrGlyPheThrThrAspGlyGlnAspTyThrLeuSerAsnGlyWe 756
4062 TATCGCCACCGCGC...CTTGATTCACAGGCTACGCGCGGCGCATTA 4108
|||:::|
756 tarGAlaAspValGlyAsnThrArgLleuArgLysAlaGlyAlaGlyTha 773
4109 AGCAGATTTATTCATCAACCGCGGCAACACATTTTCATCAGCGCTTAT 4158
|||:::|
773 lalValSerTyHisMetAspLeuGlnAsnGlyThrThrLeuGluProTrp 789
4159 TTGAGCCTGCTTATACCGATCCGCTTGGGCAAGTCCGAACAGCGCT 4208
|||:::|
790 LeuLysAlaLalValAlaArgGlnGlyLysAlaAspSerAsnGlnValLysVa 806
4209 CAATACCGCGCTATTGGCTTCAGGATTTGCGCAAAACCGGAGTGGCGAAT 4258
|||:::|
806 lAsn.....AspAspGlyLysPheAsnAspValAla 817
4259 GGGCGCTTAAC.....GCCGAAATCAAGGTTTC..... 4287

```

|||||
817 laglythSerglyValTyrGlnAlaQlyLeaIrgSerSerPheThrPro 833
4288 ACGGTGTCCTCCACGCTGCGCGCGCGCAAGCGCCGCACTGGAACGCA 4337
|||||
834 ThrleuSerglyHisLeuSerValSerTyrGly..... 844
4338 ACACAGCGCGGCGCATCAA 4356
|||||
845 .AsnGlyAlaGlyValGlu 850

seq_name: SwissProt_40:YDBA_ECOLI

seq_documentation_block:
ID YDBA_ECOLI STANDARD; PRF; 2003 AA.
AC P33666; P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ydbA.
YDBA OR B1401/B1405.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
RA Takeuchi K., Takeuchi Y., Wada C., Yamamoto Y., Horiiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN
RP SEQUENCE OF 464-2003 FROM N.A.
RC STRAIN-K12;
RX MEDLINE=92190338; PubMed=1665988;
RA Moszer I., Glaser P., Danchin A.;
RT "Multiple IS insertion sequences near the replication terminus in
Escherichia coli K-12.";
RL Biochimie 73:1361-1374(1991).
CC -1- SIMILARITY: TO S. TYPHIMURUM ONE NEAR CYS6 (AC P23928).
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
BETWEEN AMINO ACIDS 839 AND 840.
CC
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000237; AAC74483.1; ALT_SEQ.
CC EMBL: AE000237; AAC74487.1; ALT_SEQ.
CC EMBL: D90778; BAA15009.1; ALT_SEQ.
CC EMBL: D90778; BAA18880.1; ALT_SEQ.

```

```

DR EMBL: D90779; BAA18881.1; ALT_SEQ.
DR EMBL: X62680; -; NOT_ANNOTATED_CDS.
DR Ecogene; EG11307; ydbA.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 489 489 I -> V (IN REF. 2).
FT CONFLICT 495 495 I -> V (IN REF. 2).
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;

alignment_scores:
Quality: 252.00 Length: 1864
Ratio: 0.311 Gaps: 89
Percent Similarity: 43.401 Percent Identity: 19.152

alignment_block:
US-09-303-518D-649 x YDBA_ECOLI
Align seg 1/1 to: YDBA_ECOLI from: 1 to: 2003

172 GCCGAATAATTAAGCAAGTTTGCAGTCGGCGCAAGAT.....AT 212
|||||
324 AlaAsnAsnAsnGlySerThrValAspGlyLysAspSerThrClyTh 340
213 TGAGGTTTACAAACAAAAGCGAGTGTGC..... 243
|||||
340 rGluIleAsnGlyAsnAsnGlyValIleGlnAspGlyLysLeuAsp 357
244 .....GGCAATTCATGACA... 258
357 alserGlyGlyHisGlyIleAspIleThrGlyAspSerAlaThrVal 373
259 .....AAGCCCCGATGATT..... 273
374 AspAsnLysGlyThrMetThrValThrAspProGluSerIleGlyIle 390
274 .....GATTTCTGCTGTCGGCTAAGCGCGCGCATTCG 313
390 nValAspGlyAspGlnAlaValAlaValAsnGluGlySerAlaIle 407
314 TG.....GGCGATCATATATATGTGAGC 336
407 hrAsnGlyGlyThrGlyThrGlnIleAsnGlyAspAsp.....Ala 420
337 GTGGACATTAAGCGGCTATTAACACCTTATTTGTCGGCAGAG 386
421 ThrAlaAsnAsnGlyLysThrThrValAsp...GlyLysAspSerTh 436
387 AATCCGATCAACATCGTTTACTTATTAATATGTAACGGAATAT 436
436 rGlyThrGluIleAlaGlyAsnAsnGlyLysValIleGlnAspGly 453
437 ATAAAGCAGGCACTAAGCCAT.....CCTATGCGCGCATATCAT 480
453 euAspValSerGlyGlyHisGlyIleAspIleThrGlyAspSerAla 469
481 ATGCGCGCTTGCAATTAATTTGCACAGATGCAACACCTTGAATGAC 530
470 ThrValAspAsnLysGlyThrMetThrValThrAspProGluSerIle 486
531 CAGTTATATGATGCGCGCAATATATATGCAAAATATATACCTGACC 580
486 yIleGlnIleAspGlyAspGlnAlaIleValAsnAsnGluGlySer 503
581 GTGTT...CGTATGGGGCAGCGCAGCAATATTTGGCATGTGAAGAT 627
503 hrIleThrAsnGlyGlyThrGlyThrGlnIleAsnGlyLysAspAlaThr 519
628 GAGCCCAATAACCGCAAGTCAATATCATATGCAAGTGCAGTCTCTG 677
520 AlaAsnAsnSerGlyLysThrThrValAspGlyLysAspSerThrCly 536
678 GCTGTGTGGCGCAATACCTTTGCACAAATGATGAGTGTGCGACAG 727
|||||

```

536 rlysllealaglyasnilleglyllevalasnleuaspglyserleuthv 553
728 TCACATTAGTAGTGAATAAATTAACATAGCCCATATGTTTTCACCA 777
553 althreclglyalahisglyvalglu..... 561
778 ACAGAGGCTCATTTGGCGACAGTGGCTCA.....CCAA 812
562Asnilleglyaspasnnglythrvalasnasnlysglyaspil 575
813 GTTATATATAGTCCCAAAAGCAAAAGTGTATATATATG..... 855
575 evalvalseraspthrelyserilleglyvallelleasnnglyglu 592
856GTATTCGAACGGCGACCCCTATATA..... 882
592 lathrvalserasnthrclglyaspvalasnvalserasnnglylathrcl 608
882 882
609 PheSerIleThrThrAsnSerGlyLyValSerLeuAlaGlySerMetG 625
883GCAAAAAGCAATGGCT 898
625 nvalglyaspheSerThrGlyValAspLeuAsnGlyYasnAsnAsnSerV 642
899 TCCAGCTGGTTCGTAAGATTGGTTATGATGAATC..... 936
642 althrleualalalalysaspleu.....LysValValGlyGlnLys 655
937TTTGTGAGATACCCATCAGTA..... 960
656 AlathrGlyIleasnValSerGlyAspAlasnThrValAsnIleThrG 672
961T 961
672 yasnValleuValAspLysAspLysThrAlaAspAsnAlaLaGlyLyr 689
962 TCTAGCAACGCTCAAAATGGAATATCTTTTAACGAGATATAT 1011
689 hepheaspro...SerValGlyIleasnValTYrGlySerAspAsn 704
1012 GGCACA.....GCAAAATCATGCCCCAATGACACAAATCTCTGCC 1055
705 ValThrLeuAspGlyLysLeuThrValValSerasp..... 716
1056 TAATAGATTAATAACGACGACCGTCAATGTTATATCTTCTTATCCG 1105
717 .sergluValThrSerArgLysSerAsnLeuPheAspGlySerAlaGlu 733
1106 AGACGAGCAAGACGACCTGTTATCATGCTGCAGGTGCTCAAC..... 1149
733 ysrThrSerGlyLeuValValIleGlyAspGlyAsnThrValAsnMetAsn 749
1150AGTTATCGACCCAGCTGATTAATGAGAAA 1181
750 GlyGlyLeuGluLeuIleGlyGluLysAsnAlaLeuAlaAspGlySerG 766
1182 TATTTCTTATTTGACGAAAGAAAGC.....GAATTGATCTTACCA 1225
766 nvalThrSerLeuArgThrGlyTYrSerTYrThrSerValIleValValS 783
1226 GCAACATCAATCAAGTGCTGAGAGTATATTTCCAGAGATTTTACG 1275
783 er.....GlyGluSerSerValTYrLeuAsnGlyAspThrThr 795
1276 GTCCTGCCCTGAAAATAACGAAACTGGCAGGC..... 1308
796 IleSerGlyGluPheProLeuGlyPheAlaGlyValIleArgValGlnAs 812
1309GCGGCGCTTCATATCAGTGAAGACGATCCGTTACTTGGAAAGTAA 1354
812 pLysAlaLeuLeuGluIleGlySerGlyValAlaThrLeuThrMetGln.... 827
1355 ACGGCTGGCAACGACCGCTGTCCAAATGCGCAAAAGGCACGCTGCAC 1404
828AspIleAspSerPheGlyIleHisIleGlyThrArgThrValGlu 841
1405 GTTCA.....GCCAAAGGGGAAAAACCAAGCTCGAT 1436
842 IleGlnAsnLeuGlyPheAlaPheValThrGlyGlnAsn.....ThrTh 856
1437 CAGCTGGCGCAGCTACAGTCATTTTGGATCAGCAGCAGCAAGATTAAG 1486
856 rGlyIleAsnSerGlyThrIleSerLeuGln.....AsnG 869
1487 GCAAAAACAAAGCTTTAGTGAATC..... 1512
869 LyLysAspProAlaProSerProIleValIleLeuAlaThrAsnGlyGly 885
1513GCTTGTCCAGCGCGACGGGTACGCTGCACATCA 1547
886 SerAlaThrAsnAlaGlyThrIleThrGlyLysValThrGluGlnHisE 902
1548 TCCCATATATCAGTTCACACCC...GACAACTCATTTCCGCTTCCG 1594
902 rValPheAsnLysTYrSerThrGlyThrSerAsnSerPheIlePheAsnA 919
1595 GCGCAGCTTTGGATTAAACGGG...CATTCGCTTTCGTTCCACCGTATT 1641
919 snAspValSerSerIleThrGlyLeuValAlaGlnSerAsnSerThrIle 935
1642 CAAATACCGATGAGAGGGCGATG..... 1665
936 IleAsnThrAspSerGlyIleIleAspLeuTYrGlyArgGlySerValG 952
1666ATTGTCAACCAATCAAGCAAGCAAGATCCACGTTTACCA 1705
952 yMetLeuAlaIleAlaAspSerThrAlaGluAsnGlnGlyLysIleThrL 969
1706 TT.....ACGGCAATTAAGTATT 1725
969 euAspSerMetTrpValAspAlaAsnAspThrThrAlaMetArgAspIle 985
1726 GCTACA.....ACCGCAATTAACAGACGTTGGA 1754
986 AlaSerAsnSerAlaIleAspPheGlyThrGlyValGlyValGlyThrAs 1002
1755 TACG.....AAAAAGAAATTCCTCAAC..... 1779
1002 pSerTYrSerGlyAlaGlyLysAsnAlaThrAlaIleAsnGlnLeuGlyG 1019
1780GGTTGGTTGGCGAG 1794
1019 lyValIleThrIleTYrAsnAlaGlyAlaGlyMetAlaIleArgGlyAla 1035
1795 AAAGATACGACAAAGCAAGCGCGCTCAACCTTGT..... 1833
1036 SerAsnThrValIleAsnGlnGlyThrIleAsnLeuGlnLysAsnGlyAs 1052
1834 .TACACGCGCGCGCAGACAGCCACCTG.....CTGCTTT 1870
1052 nTYrAspAspSerLeuAlaIleAsnThrIleuValGlyMetAlaValTYrG 1069
1871 CCGCGGGAACAATTTAAAGGC.....AACATCAGCA 1905
1069 IuhIstGlyThrAlaIleAsnAspGlnThrGlyValIleAsnIleAsnVal 1085
1906 ACAACGCAAACTGTTTTCAGC...GGCAGACCAACA..... 1941
1086 GlyThrGlyGlnAlaPheTYrAsnAspGlyThrGlyThrIleValAsnTY 1102
1942CCGACGCTACATC 1957
1102 rGlyThrIleCYsrThrPheGlyValGlyGlnSerGlyAsnGlnLysrAsnA 1119

```
1958 ATTTAAAGCAGCATTTGGTCGCAAAAAGAGGCAATTCCTCGCGGGAAATC 2007
      : : : : :
1119 snthraspspnhetrserleu.....lle 1127
2008 GTGTGGGACAGCATGTGATCAACCCACATTAAAGCGGAACCTTCCA 2057
      : : : : :
1128 TyrtHglYglYAspThrIlethrArgSer.....GlyglutHrYalrh 1142
2058 AATTAAAGCGGACAGCGGTGTTCGCCAATGTTGCCAAAGTGAAG 2107
      : : : : :
1142 rleuAsnlysserAlaIalValIthrAspIysleuAlaGlyAsnValAla 1159
2108 GCGATTGGCATTGTGACCAATCAGCCCAAGCAGTTTGGTTCGACCG 2157
      : : : : :
1159 snSerIythrIleuSerGlyAspGlnIleThrValSer..... 1171
2158 CATCAAGCCACACATCTGTACAGCTTGAGCTGAGCGGTCTGCANA 2207
      : : : : :
1172 .....SerGlyleuIleuG 1176
2208 TTGTGTGAAAAAACCATTTACCGAGCATTAAGTGAATGCTTCATTGACTA 2257
      : : : : :
1176 u.....AsnThrSerGlyIleIleAsnAsnIleuValI 1188
2258 AGACGAC.....ATCAGCGGCAAT 2277
      : : : : :
1188 yslAuAspIysGlyAlaValIleIysAsnAlaGlyAlaIleThrAsn 1204
2278 GTGCAATCTTGGCCGATCAGCGCTCATTTAAATCTCACAGGCGTTCACACT 2327
      : : : : :
1205 ValAspValSerGly...GlyIleleuAsnAlaGlygluMetThrAl 1220
2328 CAACGGCAATCTTACGCAATAGGCAATACAGTTATACAGTACCCACA 2377
      : : : : :
1220 aGlnIleThrMetAsnAlaGlyAlaAspSerSer...leuValAsnAnt 1236
2378 ACGCCCAAAAAGGCAACCTTACGCTTCGTGGGCAATGCCACACACA 2427
      : : : : :
1236 hGlyIthrIleAsn.....LysIleValGlnAsnAlaGlyAlaIle 1249
2428 TTTAATCAAGCCACATTAACGCGCAACATCGGCTTCGGCAATGCTTC 2477
      : : : : :
1250 AsnAsnSerGlySerValIthrGlyArgMetSerIleAlaGlyIValAla 1266
2478 ATTATATCAAGCCACAGCGCGTCAAAAGCGCAGTGTGACGCTTCCG 2527
      : : : : :
1266 eaSnAsnGlnThrAspGlyAlaIleMetArgGlyAla...AlaIleuThrG 1282
2528 GCAACGCTAAGCA..... 2541
      : : : : :
1282 lyThrAlaValAlaAsnAsnGlnGlyIthrTPAsnleuGlySerSerSer 1298
2542 .....AACGTAAAGCATTCGCGACTCAACGCGTAATGTCTCCCTGAC 2582
      : : : : :
1299 GlnGlyAsnAsnThrGlyMetleuGlnValAlaAsnAsnSerAlaIleAs 1315
2583 CGATTAAGCGCATTTCCATTTTGAAGACAGCCGCTTTACCGACAAATCA 2632
      : : : : :
1315 nAsnArgIyGlnupheIleuAspAsnAspLys..... 1326
2633 GCGGCGGACAGATACGCGATTACCTTAAGACAGAGCATGAGCGCTG 2682
      : : : : :
1327 .....AsnAlaValIleHisIleAsnGln..... 1333
2683 CCGTCAGCAGC.....GAATTAGGCAATTTAAACCTTGACAAAGCCAC 2726
      : : : : :
1334 ...SerGlyIthrLeuTyAsnThrGlyHisMetAsnIleSerAsnSerSe 1349
2727 C..... 2727
      : : : : :
1349 rHisAsnGlyAlaValAsnMetTPGlyGlyAsnGlyArgpHeIleAsnA 1366
2728 .....ATTACACTCAATTC 2742
```

```
1366 spGlyThrIleAspValSerAlaIySerleuValIalSerAlaAsn 1382
      : : : : :
2743 GCC..... 2745
      : : : : :
1383 AlaGlyAspGlnAsnAlaIlePheThrAsnGlnAspAsnGlyAlaIleAs 1399
2746 .TATCGCACATGCGCA..... 2763
      : : : : :
1399 nPheAspHisAspSerAlaSerAlaValAlaIysValIthrHisSerAsnPheI 1416
2764 .....GGGCGCAACCGGCAAGTGGCGAGCATGGCGCGC 2799
      : : : : :
1416 leaIaGlnAsnAspGlyIleMetAsnIleSerGlyIthrGlyAlaValAla 1432
2800 GCGCGTTCGCGCGGTTCGCGCGCTTCCTATATATCCGTTACACCGGCCAC 2849
      : : : : :
1433 MetGlnGlyAspIysAsnAlaGlnleuValAlaAsnAsnGlyIthrIleAsnle 1449
2850 TTCGGTGAATCCCGTTTCAACACAGTGCAGCGTAAACGGCAATTGACG 2899
      : : : : :
1449 uGlyThrAlaGlyIthrThrAspIthrGlyMetIleGlyMetGlnleuAspA 1466
2900 GTCAAGACATTCGCTTATGTGGGAACCTTCGCGC..... 2937
      : : : : :
1466 laAsnAlaIthrAlaIaspaIalValIleGlnAsnAsnGlyIthrIleAsnIle 1482
2938 TACCGCAGCCGCAAAATTGAAGCTGGCGGAAGTTCCGACGACCTTACAC 2987
      : : : : :
1483 PheAlaAsnAsp.....SerPheAlaIlePheSerValIleuGlyIthrValG 1497
2988 CTTGCGCGTCAACAT..... 3003
      : : : : :
1497 yHisValAlaIAsnAsnGlyIthrValValIleAlaAspGlyValIthrGly 1514
3004 .....ACCGC 3009
      : : : : :
1514 ergIleuIleIysGlnGlyAspSerIleAsnValGlnGlyMetAsnGly 1530
3010 AAGCAACCTGCAGCCTCGAA..... 3030
      : : : : :
1531 AsnAsnGlnIysAsnSerGlnValIleHisIyIAspTyIthrLeuProAs 1547
3031 .....CAATTGACGAGTGTGAAGCAAAAGCAACAAC 3064
      : : : : :
1547 pValIProIyProAsnIthrValSerValIthrSerGlySerAspGlnAlaG 1564
3065 CGCTGTCC...GAACCTTAATTTCACCTGCAGAAACGACGTCGAT 3111
      : : : : :
1564 lyGlySerMetAsnAsnleuAsnGlyTyIValAlaGlyIthrAsnValAsn 1580
3112 GCGCGCCGTGGGCTTACCAACATCCGCAAGACGGGAGTCCGCCCT 3161
      : : : : :
1581 GlySerIleA.....GlyIysleuIysVal 1588
      : : : : :
3162 GCATTAAT...CCGTCAAAGACAGAGCTTTCGCAAACTCGGCAAG 3208
      : : : : :
1588 laAsnAsnAlaSerMetAsnGlyValGlnIleAsnIthrGlyIthrAlaG 1605
3209 CAGAAAGCAAAAACAGCGCGGAAAAAGACACCGCGCAAAAGCTTGACGCG 3258
      : : : : :
1605 lyThrAlaAspIthrIthrValSerPheAspAsnValAlaGlnGlySerAsn 1621
3259 CTGATTGCGCGCGCGCGATGCGCTTCGAAAGCAAGAAAGGTT..... 3303
      : : : : :
1622 leuThrAspAla.....AspAlaIleIthrSerIthrSerValIalIthrPh 1636
3304 .GCCGAACGGCGCGCGCAGCGAGCGGGGAAATGTGCGCATTAATGACG 3352
      : : : : :
1636 rAlaIysGlySerIthrAspAlaSerGly...AsnValAspValIthrMetS 1652
3353 CGGAGGAAGAAAAACGGGTGCGAGCGGATTAAGACACCGCCTTGGCG 3402
```


[illegible]


```
2638 GGCAGGATACGGCATTTACTTA.....AAAGACGCAATGAC 2678
|||||
1300 GlyIuAsnThrIleIleAsnLeuThrAlaIAsnSperLeuThrGI 1316
2679 GCTCCGCTGAGCGACGGAATTAAGCAATTAACCTTGACACGCCACCA 2728
|||||
1316 nValAsnSerLeuThr...IleThrIleThrAsnGlyTyrSerAlaIleI 1332
2729 TTACATCAATTCGGC.....TATGCCACGATGGCGCA 2763
|||||
1332 hrThrValAsnGlySerAsnSperProLysTrpIleThrAsnThrAlaThr 1348
2764 GGGCGCAAAACGGCGACGTGGACAGATGCCGCCGCCGCTGGCGCGC 2813
|||||
1349 GlyValIleAsnGlyIleAsnProAspAlaPro..... 1359
2814 TTCCGGCGCTCCCTATATTCGTTACCGCCACTTGCGTAAGATCC 2863
|||||
1360 .....LeuIleAsnLeu.....GlyA 1365
2864 GTTTCACACGCTGACGCTGAACGCAAAATGAACGTCAGGAAACATTG 2913
|||||
1365 rGgIyTyraSnpheGlyAsnGlnGlyThrIleAsnValGlnGly..... 1379
2914 CGCTTATGTGCACTCTTCGCTACCGCAGCAAAATGAAGCTGGC 2963
|||||
1380 .....AspAsnAlaValAlaIle 1385
2964 GGAAGTTCGGAAGCACTTACCTTGGCGGTCACCAATACCGGCACG 3013
|||||
1385 eSerGlyGlyThrSerSerTyrValIleAsnLeuValAsnSerGlyThrI 1402
3014 AACCTGCAGCCTCGAACA..... 3033
1402 leAsnValGlyThrGlnGlnGlyLysGluAspGlyThrAsnGlyThrGly 1418
3034 TTGACGGTAGTGGAAGAAAGCAAAACAGCCGTGCGAAACCTTAA 3083
|||||
1419 LeuIleGlyIleLysGlyAsnGlyAsnAlaThrThrIleAsnAspThrAl 1435
3084 TTTACCCCTGCAGAAAGCAACAGCTGCAT.....GCCGGCGGCTGGC 3124
|||||
1435 aaSpGlyValIleAsnValTyrAlaAspAspSerTyrAlaPheGlyLyl 1452
3125 GTTACCACTCATCCGCAAGAGCGGAGTTCGCGCTG..... 3162
|||||
1452 ysthrLysAlaIleIleAsnAsnGlyIleAsnLeuLeuCyaspser 1468
3163 .....CATAA 3167
1469 GlyCyaspIleTyrAlaProGlyThrThrGlnAsnAspHisAs 1485
3168 T..... 3168
1485 nGlyThrAlaAspIleValIleProAspAlaThrThrAlaProThrGluG 1502
3169 .....CCGTCAGAAAGCAAGACCTTTC 3192
1502 ySerIleProThrProAlaAspProAsnAlaProGlnGlnLeuSer 1518
3193 GACAAACTC.....GGCAAGCGAAGC 3215
|||||
1519 AsnTyrIleValGlyThrAsnAlaAspGlySerSerGlyThrLeuLysAl 1535
3216 CAAAAAAGCGCGGAAAAAGACAAAGCGCAAGCTTGACGCGCTGATTG 3265
|||||
1535 aAsnAsnLeuValIleGlyAspAsnValLysValAspThrGlyPheThrS 1552
3266 GCGCGGCGCGGATGCGCGTGAAGAAAGCAAGACGTTGCCGACCGCC 3315
|||||
1552 ergIyThrAlaAspThrThrValValAlaValAspAsnAlaPheThrGlySer 1568

3316 CGCAGCGAGCGCGGAAAAATGTC.....GGCATTTACGAGCGGA 3356
|||||
1569 AsnIleGlnGlyAlaAspAsnIleThrSerThrSerValValTrpAsnAl 1585
3357 GGAAGAGAAAAAGCGGTGACAGCGCGATTAAGACACCGCTTGCGCAAC 3406
|||||
1585 aGlnGlySerIleAspAlaAspGlyAsnValAspValThrMetThrLysA 1602
3407 AGCGCAAGCGGAAACCGCGCGGTACACCGCTTCCCGCGCGCCG 3456
|||||
1602 snAlaTyrAlaAspValAlaThrAspSerSerVal..... 1613
3457 GCGCGCGCGCGGATTTGCGCACTGCAACCCCAACCGACCGCCAC 3506
|||||
1614 .....SerAspValAlaGlnAla..... 1619
3507 GCACGCGCGATGATCAGCGCTTATGCCAATAGCGTTG..... 3546
|||||
1620 .....LeuAspAlaGlyTyrThrAsnAsnGlnLeuTyrThrSerL 1633
3547 .....AGTAATTTTCCGCAAGCGCTGCAACGCGTTTC 3579
|||||
1633 euAsnValGlyThrThrAlaGlnLeuAsnSerAlaLeuLysGlnValSer 1649
3580 GCGGTACAGGACGAATTAAGACCGCATTTGCGCAAGACCGC..... 3621
|||||
1650 GlyAlaGln.....AlaThrThrValPheArgGluAlaArgValLeuSe 1664
3621 ..... 3621
1664 rAsnArgPheThrMetLeuAlaAspAlaAlaProGlnIleLysAspGlyL 1681
3622 .....CGCAACGCGCTTGGACACGCGCATCCG..... 3651
1681 euAlaPheAsnValAlaAlaLysGlyAspProArgAlaGlnLeuGlyAsn 1697
3652 GACACCAACACACCTTCGCAAGATTTCGCGGCTGACCGCAACAAAC 3701
|||||
1698 AspThrGlnThr.....AspMetLeuAlaLeuArgGlnThrLe 1710
3702 CGACCTGCC.....CAATTCGTAATGCAAGAA 3730
|||||
1710 uAspLeuThrAlaSerGlnAsnLeuThrLeuGlnLysGlyLeuAlaArgL 1727
3731 ACCTGGCGGCGCGCGCTCGGCATCTGTTTTCGCAACCGACCGAA 3780
|||||
1727 euAspLysAspLys.....SerLysThrAlaGlyAsp 1737
3781 AACACCTTGCAGCAGCGCATCGCAACTGGCAGCGCTTGCCGACGGCGC 3830
|||||
1738 AsnGlyLeuThrGlyGlyTyrSerGlnPhePheGlyLeuLysHisSerMe 1754
3831 CGTTTTGGGCA.....T 3844
1754 tAlaPheAspGlnGlyLeuAlaTrpAsnAsnSerLeuArgTyrAspValH 1771
3845 ACGGCATGACAGGTTTACATCGGCATCAGCGCGCGCGGCTTTTACG 3894
|||||
1771 tAsnLeuAsp.....Ser 1775
3895 AGCGGACGCTTTCAGACGCGCATCGAGCAAAATC..... 3930
|||||
1776 SerArgSerValAlaTyrGlyAspValAsnLysIleAlaSperSperme 1792
3931 GCGCGCGCGCTGCATTC.....GGCATTCAGCAGCATTAC 3970
|||||
1792 tArgGlnIleTyrLeuGlnPheArgSerGlnGlyAlaLysThrPheThrM 1809
3971 GCGCGGCTTTCGCGATTCGCAACCGCAACGCAATCGCGCAAGCGCC 4020
|||||
1809 etMetGly.....AspAlaLeuLysValThrProTyrAlaGlyValLys... 1823
4021 TATTCGTCAAAAAGCGGATTAACCGCTACGAAAGCAATATGCGCAC 4070
```



```

281 1A.....AlaIaIThrAspPheAsnValThrGlySerLeuGlyGly 294
946 GATTAACCATTCAGTA.....TTCTACGAAACACAGCTCAAAATGGGAATA 999
295 AsnLeuLysGlyIleIleGluPheAsnThrValAlaValAsnGlyGlnIle 311
990 CTCTTTTAA.....GACGATATAATGGCA 1015
311 ULysAlaAsnAlaGlyAlaAsnAlaIleValIleGlyThrAsnAsnGlyA 328
1016 CAGGAAAAATCAATGCCAACATGACACAAATTCCTGCTTAAT...AGA 1062
328 IaGlyArg.....AlaIaGlyPheValIalSerValaIaAspAsnGlyLys 342
1063 TTAATAACAGCAACCGTTCAATGTTTAAATGTTCTTATACGACAGAC 1112
343 ValAlaThrIleAspGlyGlnValTyr.....Al 352
1113 AAGGAACCTGTTTATCATGCTGCAGGTGTCAC.....AGTTATC 1156
352 aLysAspMetValIleGlnSerAlaAsnAlaValIleGlyGlnValaIaAsnPheA 369
1157 GACCCAGACTGAATATGGA...GAAATATTCCTTTATTCAGCAGGA 1203
369 rGHisIleValaIspValIleGlyThrAspGlyThrIaPheLysThrAla 385
1204 AAGGCGAATGATTAATACAGCAACATCAAGTGT..... 1245
386 AlaserLysValAlaIleThrGlnAsnSerAsnPheGlyThrThrAspH 402
1246 GGAGGATTTATTTCCAA.....G 1264
402 eGlyAsnLeuAlaIleGlnIleValProAsnThrMetThrLeuAsnG 419
1265 GAGATTTTACGTCGCTGGAATAAGCAAACTTGCAAGCGCGCGG 1314
419 LysAsnPheThrGlyAspAlaSerAsnProGlyAsnThrAlaGly..... 433
1315 GTTCATATCATGTAAGACAGTACCGTTACTTGAAAGTAAAGGC..... 1359
434 .....ValIleThrPheAspAlaAsnGlyThrIle 443
1360 .....GTGCAACACGCGCGCTGT 1378
443 uAlaSerAlaSerAlaAspAlaAsnValAlaValThrAsnAsnIleThrA 460
1379 CCAAAATCGCAAGGACGCTGCAGCTTCAAGCCAAAGGGGA..... 1422
460 IaIleGluIaIaSerGlyAlaGlyValaIleGlnLeuSerGlyThrHisAla 476
1423 .....AACCAAGCTTCGATC...ACGTTGGCGGACGG 1451
477 AlaGluLeuArgLeuGlyAsnAlaGlySerValPheLysLeuAlaAspG 493
1452 TACAGCTATTTTGATTCAGCAGCAGAGATAAGTAAACAAACAGGCT 1501
493 yThrValIle.....AsnGlyLysValaIaAsnGlnT 503
1502 TTAGTGAATCGGCTTGGTCAGCGGACGGGTACGGTCACTGAATGCC 1551
503 hAlaIaLeuValGlyIleValaIaLeuAlaIaGlyThrIleThrLeuAspGly 519
1552 GATTAATCAGTTCAACCCGACAAACTCATTTGGGCTTCCGGCGGACG 1601
520 SerAlaThrIleThrGlyAsp.....IleGlyAsnAlaGlyGlyAl 533
1602 TTTGATTTTAAAGCGCATTCGCTTTCGTTCCACCGTATTCAAAT...A 1648
533 aAlaIaLeuGlnGlyIleThrLeuAlaAsnAspAlaThrLysThrLeuT 550
1649 CCGATGAAGGCGCATGATGTCAAACCAAT..... 1680
550 hLeuGlyGlyAlaAsnIleIleGlyAlaAsnGlyGlyThrIleAsnPhe 566
1681 CAAGCAAGAAGATCCACCGTTACCATTAACGACGCAATTAAGT..... 1722
567 GlnAlaAsnGlyGlyThrIleLysLeuThrSerThrGlnAsnIleVala 583
1723 .....ATTGCTACA.....ACCGGC.....ATA 1741
583 lValaIspPheAspLeuAlaIleAlaIleThrAspGlnThrGlyValaIaAspA 600
1742 ACAACAGCTGGGATAGCAAAAAAGAAATTCCTACACGCTTGTTGGC 1791
600 lAserSerLeuThrAsnAlaGlnThrLeuThrIleAsnGlyLysIleGly 616
1792 GAGAAAGAT...ACGACCAAAACGAGCGGCTCAACCTGTTTACCA 1838
617 ThrValGlyAlaAsnAsnLysThrLeuGlyGlnPheAsnIle..... 630
1839 GCCCGCGGCAAGAGACCGCACCTGCTTCCTCCGGCGGACAAATTTAA 1888
631 .....GlySerSerLysThrValLeuSerAspGlyAspValAlaIleA 645
1889 ACGGCAACATCAGCAACAAACGCGCAACTGTTTTCAGCGGCGAGACA 1938
645 sngIuLeuValIleGlyAsnAsnGlyAlaValaIlePhe..... 657
1939 ACACCGCACCGCTACAT...CATTTAAACGACCATTTGTCGCAAAAAGA 1985
658 .....AlaHisAsnThrLysLeuIleThrArgThrThrAsnAlaI 671
1986 GGGCATTCCTCGCGGGGAAATCGTGGGACACAGCATGATCAACCGCA 2035
671 aGly.....GlnGlyLysIleIlePheAsnProValaValaAsnAsnA 686
2036 CATTTAAAGCGGAA..... 2049
686 hThrIleuAlaThrGlyThrAsnLeuGlySerAlaThrAsnProLeuAla 702
2050 .....ACTTCCAATTTAAAGCGGACAGCGGCG.....GTTTCCGCGCA 2090
703 GluIleAsnPheGlySerLysGlyAlaIaAsnValaIspThrValLeuAs 719
2091 TGTGGCAAAAGTGAAGGCGATTGGCATTTGACAAATCACGCCCAAGAG 2140
719 nValGlyLys..... 722
2141 TTTTGGTGTGCGACCGCATCAACGACACATCTGTACAGCTTCGAC 2190
723 .....GlyValaIaLeuLysrAlaThrAsnIleThrThrAspAlaAsn 737
2191 TGG.....ACGGGTCTGACAAATTTGTGCGAAAAAC 2222
738 ValGlySerPheIlePheAsnAlaGlyGlyThrAsnIleValSerGlyThr 754
2223 CAT.....ACCGAG 2233
754 rValGlyGlyGlnGlnGlyAsnLysPheAsnThrValaIaLeuAspAsnG 771
2234 ATAAAGTATGCTCATTTAGCTAAGACCGACATCAGCGCATGTGCAT 2283
771 LysThrThrValLysPheLeuGlyAsnAlaThrPheAsnGlyAsnThrThr 787
2284 CTGGCGATCAACGCTCATTTAAATCTCACAGGCTTCGACACATCAACG 2333
788 lIleAlaIaAsnSerThrLeuGlnIleGlyLysnLysrThrAlaAspH 804
2334 CAATCTTACGCAAAATGGCGATACAGCTTATACGTACGACACACACCA 2383
804 eValaIaSerAlaAspGly..ThrGlyIleValaIleGluPheValaIaThrGly 820
2384 CCAAAAGGCGCAACCTTAGCTCGTGGGCAATGGCGAAGCAACATTTAAT 2432
821 ProIleThrValThrLeuAsnLysGlnAlaIaIaProValaIaAsnAlaLeu 837
```

| | | |
|------|--|---------------------------|
| 2433 | ..CAAGCCATTAAACGGCAACATCGCGCTTCGGCAACGCTTCAATT | 2482 |
| | | |
| 837 | sglnlrrhrvalsergylproglyasnval leasgnlileglya | 854 |
| 2483 | ATCTAGCCACAC.. | ..GCCGTACAAAC |
| 2508 | 2483 | 2508 |
| 854 | snlactylasnrrhrslsglyalaValThrAspThrIleAlaPheGluAsn | 870 |
| | | |
| 2509 | GGCAGTCTGACGCTT | 2523 |
| | | |
| 871 | SerSerleuglyalaValAlaPheLeuProArgglyleProPheAsnAs | 887 |
| 2524 | ..TCGGCAAC.. | ..GCTACGAAACGTACGCAATTCCG |
| | | |
| 887 | palaglyasnArgyleProLeuThrIleIysSerThrValGlyAsnLysr | 904 |
| 2558 | CACCTCAACGGTAATGTCTCCCTACGCCGATTAAGCAGATATTCATTTTGA | 2607 |
| 904 | hrAlaTrnGly..PheAspValProSerValIleValLeuglyalaAsp | 919 |
| 2608 | AGCAGCCGCTTACCGCAAAATTCAGCGCGCAAGATACGCAATTACA | 2657 |
| | | |
| 920 | SerValIleAlaAspIylnValIleIleIysAspIasnAsnIleValGl | 936 |
| 2658 | CTTA.....AAAGACAGCA.. | ..TGACGCTGC |
| | | |
| 936 | yleuglyleuglySerAspAsnAspIleIleValAsnAlaThrThrLeuTr | 953 |
| 2684 | CGTCGACGACGAAATTAAGCAATTAACCTTGACAAACGCCACATTACA | 2733 |
| | | |
| 953 | ylAlaIgly.....IleIglyThrIleAsnAsnAsnIleIlyThrValThr | 967 |
| 2734 | CTCAATTCGCGCTATCGCCACGATGCGGAGGGCGCAACCGCACATGC | 2783 |
| | | |
| 968 | LeuSerIglyIle...ProAsnThrProGlyThrValITyrglyleuGl | 983 |
| 2784 | GACAGATGCGCGCGCGCGCTTCGGCGCGTTCGGCGCTTCCTATTAT | 2833 |
| | | |
| 983 | ylThrIglyIleIyAlaSerIysPheIysGln..... | 993 |
| 2834 | CGCTTACACGCCCAACTTCGCTGATGATCCGTTTCAACAGCCTGACGGTA | 2883 |
| | | |
| 994 | ..ValThrPheThrThrAspTyraAsnAsnIleIleAlaThr | 1009 |
| 2884 | AAC..... | 2886 |
| | | |
| 1010 | AsnAlaThrIleAsnAspIyAlaThrValThrThrglyIleAlaGl | 1026 |
| 2887 |GCAAAATGAAACGGTCAAGGAA | 2908 |
| 1026 | ylIleIyPheAspIyIysIleThrIleuIySerValAsnIyAsnIyGl | 1043 |
| 2909 | CATTCCGCTTATGTGGAACCTCTCGCTACCCGACGCAACAATTGAAG | 2958 |
| | | |
| 1043 | snValIArgPheValAspIyIleIeuSerHisSerThrSerMetIleIy | 1059 |
| 2959 | CTGGCGGAAAGTCCGAAGGACACTTACCTTCGGC.....GT | 2996 |
| | | |
| 1060 | ThrTrIyAlaAsnAsnIyThrValThrTyIleuglyAsnAlaPheVa | 1076 |
| 2997 | CACAAATACCGCAACAA.....CCTGCNAACCTCGAACAATTGACGG | 3040 |
| | | |
| 1076 | IeIyAsnIleIyIysPseIserAspThrProValAlaSerValArgPheTrnG | 1093 |
| 3041 | TAGTGGAAGGAAAGACAACAACCCGTCTGCCAAAACCTTAATTTCAC | 3090 |
| | | |
| 1093 | IySerAspIy.....GlyAlaIyIeuIyIleIyAsn | 1103 |
| 3091 | CGGCAAAACGAACAGTCGATGCGCGCGCTGCCTTAC..... | 3128 |
| | | |
| 1104 | IleIyIySerGlnValIleAspPheIyThrTyraIyAsnIleuglyIleSerAs | 1120 |

```

1347 lmeFAlagly..ProgluAlaglyAlaIleProAlaAlaValAlaIaG 1363
1363 CGGATTACCGGTACGAAACGTCATATTCGCCACCCCGCTTGCAATTC 4086
1369 lYAspGlAlaValAlaAspAsnValAlaTyrGlyIleTrrAlaIAspProPhe 1379
4087 .....AACCGCTACCGCGCGCATTAAGCAGATATTCATT 4124
1380 TyrThrAspAlaHisGlnserLysLysGlyGlyLeuAlaIleTrrLysAl 1396
4125 CAACCGCGCG.....AACCGCTACCGCGCATTAAGCAGATATTCATT 4141
1396 aLysThrThrGlyValValIleGlyLeuAspThrLeuAlaAsnAspAsnL 1413
4142 TTTCATCAGCGCTTATTTGAGCTGCTATACCGCATTCGCGCTTGGCGG 4191
1413 eumETIleGlyAlaAlaIleGlyIleThrLysThrAsp..... 1425
4192 AAAGTCGGAACACCGCTCAATACCGCGCTATTCGCTAGAGATTTCGCA 4241
1426 .....IleLysHisGlnAspTrrLysLys 1433
4242 AACCCGAGTGGCGAATGGCGCTAAACGCCCAATCAAGTTTCACGC 4291
1433 S.....GlyAspLysThrAspValAlaAsnGlyPheSerP 1444
4292 TGTCCCTCAGCGCTGCGCGCGCAAGCGCGCAACTGGAGCGCAACAC 4341
1444 heserLeuTyrGlyAlaGlnGlnLeuValLysAsnPhenAlaGlnGly 1460
4342 AGCGCG 4347
1461 SerAla 1462

```

seq_name: SwissProt_40:OMPA_RICRI

seq_documentation_block:

```

ID OMPA_RICRI STANDARD: PRT: 2249 AA.
AC P15921:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein A precursor (190 kDa antigen) (cell surface
GN antigen) (rOmpA) (rOmp A).
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_taxid=783;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN:R:
RX MEDLINE=90354033; PubMed=2117566;
RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;
RT "A protective protein antigen of Rickettsia rickettsii has tandemly
RT repeated, near-identical sequences.";
RL Infect. Immun. 58:2760-2769(1990)
CC -1- FUNCTION: ELICITS PROTECTIVE IMMUNITY.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC -1- SLAYER WITH HEXAGONAL SYMMETRY.
CC -1- PTM: GLYCOSYLATED (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31227; AAA26380.1; -

```

```

DR PIR: A41477; A41477.
DR InterPro: IPR003858; romPA_rOmpB.
DR Pfam: PF02708; romPA_rOmpB; 1.
KW Antigen; Repeat; Signal; Cell wall; S-layer; glycoprotein.
FT SIGNAL 1 28
FT CHAIN 29 2249
FT DOMAIN 212 1180
FT REPEAT 212 1180
FT REPEAT 286 1180
FT REPEAT 287 358
FT REPEAT 359 430
FT REPEAT 431 505
FT REPEAT 506 577
FT REPEAT 578 724
FT REPEAT 725 799
FT REPEAT 800 874
FT REPEAT 875 949
FT REPEAT 950 1021
FT REPEAT 1022 1093
FT REPEAT 1094 1165
FT REPEAT 1166 1180
FT SEQUENCE 2249 AA; 224333 MW; A9D664C089DE087 CRC64;

```

alignment_scores: Quality: 242.50 Length: 1499
Ratio: 0.357 Gaps: 75
Percent Similarity: 45.364 Percent Identity: 20.080

alignment_block:

US-09-303-518D-649 x OMPA_RICRI

Align seg 1/1 to: OMPA_RICRI from: 1 to: 2249

```

655 CATATTGCAAGTGGCTATTCTTGGCTGTTGCT.....GGCAATACC.. 696
938 AsnLeuAsnGlyAlaLeuSerGlnValThrGlyAspIleGlyAsnThrAs 954
697 .....TTTGCACAAATATGATCAGTGGTGCGACAGTCACTAAGTACTG 742
954 nSerLeuAlaThrIleSerValGlyAlaGlyThrAlaThrLeuGlyGly 971
743 AAAAAATTAAACATAGCCCATGTGTTTATTCACAAAGAGGCTCATTT 792
971 lValAlaIleLys..... 974
793 GCGGACAGTGGCTCACCAATGTTATCTATGATGCCAAAGCAAAAGTG 842
975 .....AlaThrThrLysLeuThrAspAlaIleSerAlaValLys 988
843 GTTAATTAATGGGTATTCGCAACCGGCAAC.....CCCTATATAGAAAA 889
988 spherThrAsnProValValValThrGlyAlaIleAspAsnThrLysAsn 1005
890 GCAATGGCTTCAGCTGGTGTGTAAGATGTTCTATGATGAATCTTT 939
1005 lAsn..... 1006
940 GCTGAGATTAACCATTCATGATTCAGACACCAAGTCAAAATGGGAATA 989
1007 ..... 1010
990 CTCCTTTAAGCAGATTAATTAATGACAGAGAAATCAATGCCAAACATG 1039
1010 lThrThrGlyAsnSerThrValThrGlyAsnValGly.....A 1024
1040 AACCAATTTCTGCTCT..... 1056
1024 snThrAsnAlaLeuAlaThrValAsnValGlyAlaGlyLeuLeuGlnVal 1040
1057 .....AATAGATTAAACAGCAACGCTCAATGTTATGTTCTTTCTTT 1100
1041 GlnGlyGlyValValLysAlaAsnThrIleAsnLeuThrAspAsnLase 1057

```


1635 yserValserglyValGlyThrAlaAlaAsnSerValGlyAspIleT 1652
2402 GCCTCGTGGGC.....AATGCCCAAGCACA 2427
1652 hrThAlaGlyAlaThrSerPheAlaSerSerValAsnAlaLysGlyThr 1668
2428 TTTAATCAAGCCACATTAAAGCGACACATCGGCTTCGGGCAATGCTTC 2477
1669AlaThrLeuGlyGlyThrThrSer..... 1676
2478 ATTAAATCAAGCCACCGCCGACAAAGCGAGTGTGACGCTTTCG 2527
1677PheAlaAsnThrPheThrAsnThrGlyAlaValThrLeu.... 1689
2528 GCAACGCTAAGGCAAGCAATCGCATTCGCAACGCTAATGTCTCC 2577
1690AlaLysGlySerIleThrSerPheAlaLysAsnValThrAlaThr 1704
2578CTAGCCGATTAAGCAATTCATTTTGAAGCAGC...CGCTT 2618
1705 SerPheValAlaAsnSerAlaThrIleAsnPheSerAsnSerLeuAlaPh 1721
2619 TACCGGCAAAATCAGCGCGGCGGAGATAGCGATTACACTTAAAGACA 2668
1721 eAsnSerAsnIleThrGlyGly..... 1728
2669 GCGAATGAGCGCTGCCGTGAGCGAGCAATTAAGCAATTTAAACCTTGAC 2718
1729GlyThrThrLeu...ThrLeuGlyAlaAsn 1737
2719 AAGCCGCACTTACACTCACTTCGCTTACGCCACGATGCGGACGGGCG 2768
1738 GluValThrThr..... 1742
2769 GCAACCGCGAGTGCAGACATGCGCGCGCGCGCTTCGCGCGCTTGC 2818
1743 .GlyThrGlySerPheThrAspThr..... 1750
2819 GCGGTTCCCTATATCGGTTACACCGCCCACTTCGCTAGATCCGCTTTC 2868
1751LeuThrLeuAsnThrThrPheAspGlyAlaAlaLysSerGly 1765
2869 AACACGCTCAGGTAAACGCGCAA.....TTGAACGGTACGAGGAGC 2909
1766 AsnIleLeuIleLysSerGlySerThrLeuAsnLysSerGlyValSerTh 1782
2910 ATTCCGCTTATGTCGGAACCTTCGCTACCGCAGCAGCAAAATTTGAAGC 2959
1782 rLeuAlaLeuValValThrAlaThrAsnPheAspMetAsnAsnIle.... 1797
2960 TGGCGGAAAGTTCGAGAGCACTTACACTTCGCGGTCACAAATTC... 3006
1798SerProAspThrIleThrValIleSerAlaGluThrAla 1811
3007 ...GGCAACGAACCTGCAAGCCTGCAACAAATGACGTAAGGAAGAAA 3053
1812 GlyGlyLeuLysProThrSerLysGluAsnValLysIleThrIleAsnAs 1828
3054 AGACAACAACCGCTGTC.....GAAACCTTAATTTTCACCTGC 3094
1828 naspsAsnArgPheValAspPheThrPheAspAlaSerThrLeuThrLeuP 1845
3095 AAAACGAACACGTCGATCGCGCGCGGTTCACCAACTCATCCGAAA 3144
1845 heAlaGluAspIleAlaIleAspVal.....Ile 1854
3145 GAGCGGAGTTCGCGCTGCATATCCGCTC..... 3174
1855 AspGlyAspPheAlaProGlyGlyProLeuAlaAsnIleProAsnAlaAl 1871
3175AAGACAAAGACTTTCGACAAACTCGCAAGGAGGAG 3214

1871 AsnIleLysLysSerLeuGluLeuMetGluAspAlaProAsnGlySerA 1888
3215 CCAAAAAAGCGCGGAAAAAGAC.....AACGCGCAAGC 3249
1888 sPAlaArgIleAlaPheAsnAsnPheGlyLeuMetThrProLeuGlnGlu 1904
3250 CTGACGCGCTGATTCGCGCGCGCGGATCGGCGGAAAAAGACAAAG 3299
1905 AlaAspAlaThrThrThrIleLeuIleGlnAspValValLysProSerAspTh 1921
3300 CGTTGCCGAACCGGCGCGGACGAGCGAGGGAATGTGCGCATTTGC 3349
1921 rIleAlaAlaValAsnAsnGlnValValAlaIleAsnIle..... 1934
3350 AGCGGAGAGAAAGAAAAAGGTCGACGCGGATTAAGACACCGCTTCG 3399
1935SerSerAsnIleThrAlaLeu 1941
3400 GCGAACAACGCGGAGGAAACCGCGCGCTACACCGCGCTTCGCCG 3449
1941 1941
3450 CGCGCGCGCGCGCGCGGATTTGCGGCACTG...CAACCCCAACGC 3496
1942AsnAlaArgMetAspLysValGlnSerGlyAsnLysGlyProV 1956
3497 AGCCCAACCGCGAGCGGACCTGATCAGCGCTTATGCCAATGACGCTTG 3546
1956 alSerSerGlyAspGluAspMetAspAlaLysPhe...GlyAlaThrIle 1971
3547 AGTGAATTT...TCCGCCACG.....CTCAACAGCGTTTTCGC 3581
1972 SerProPheValGlyAsnAlaThrGlnLysMetCysAsnSerIleSerG 1988
3582 GTTACAGACGAATTAACCGCGGCTATTTCGGAAGACCGCGCAACGCGC 3631
1988 yTryLysSerAspThr..... 1993
3632 TTGGACAGCGGACCTCGGAGACACCAACACTACCTTCGCAAGATTTC 3681
1994ThrGlyThrIleGlyPheAspGlyPheValSerAspAspLeu 2008
3682 CGC.....GCTTAC...CGCCCAACAACGACCTGCGCCCAATTCG 3719
2009 AlaLeuGlyLeuAlaThrThrArgAlaAspThrIleLys..... 2022
3720 TATGCAAAAAACCTCGGACGCGCGCGCTGCGCATCTGTTTTCGACA 3769
2023LeuLysAsnA 2026
3770 ACCGAGCGGAAACACCTTCGACGACGACATCGGCAACTCGGACGCGCT 3819
2026 snLysThr.....GlyAspLysAsnLysValGluSer 2036
3820 GCCCAACGCGCGCTTTCGGCAATPACGACATCGACAGGTTTCATTCG 3869
2037 AsnIleTySerLeuTyGlyLeuTyAsnValProTyGluAsnLeuPh 2053
3870 CATCAGCGCGGCGGCTTTCAGCAGCGGAGCCTTCAGACGCGCATCG 3919
2053 eValGluAlaIleAlaSerTySerAsp..... 2062
3920 GAGGCAAAATCCGC.....CGCGCGTGTGCTATTCAGGCAATTCAG 3960
2063 ...AsnLysIleArgSerLysSerArgValIleAlaThrThrLeuGln 2078
3961 ...GCACGATACCGCGCGGTTTCGGGATTCGCGCATTCGACCCGACAT 4007
2079 ThrValGlyTyGlnThrAlaAsnGlyLysTyLysSerGlnSerTyThr 2095
4008 CGGCGCAACGCGCTATTCGTCCAAAAAGCGGATTCACCGCTAC..... 4050
2095 rGlyGlnLeu.....MetaGlyTyThrTyMetMeCP 2107

4051 ..GAANAGCTCATATGCGACCCCGGCTTCATTCACCGCTACGCG 4098
 |||||..... ||||| ||||| |||||
 2107 roGlnsnlleasnleu...trpPro...LeuAlaGlyLeuArgTyr... 2120
 4099 GCGGCGCATTAAGCGAGATTATTCATCAACCGGCG.....CA 4136
 ||||| ||||| ||||| ||||| ||||| |||||
 2121 SerThrIleLys...AspLysGlyTyrLysGluThrGlyThrTyrG1 2136
 4137 ACACATTTCATCAGCCCTTATTGACGCTGCTCATACCGATGCGGTT 4186
 ||||| ||||| ||||| ||||| ||||| |||||
 2136 naenleuthrValLysGlyLysasnTyrasnThrPheaspGlyLeuLug 2153
 4187 CCGGCAAGTCCGACACGCGTCAATAC..... 4215
 ||||| ||||| ||||| ||||| ||||| |||||
 2153 LysAlaLysValSerSerAsnIleAsnValAsnGluLeuValLeuThrPro 2169
 4216GCCGTAFTGGCTCAGATTTCGCGCAACCCGACGTCGGA 4256
 ||||| ||||| ||||| ||||| ||||| |||||
 2170 GluLeuTyrAlaMetValAspTyrAlaPheLysAsnLysValSerAla.. 2185
 4257 ATGGGCGGTAAAGCGCAATCAAGGTTTCAGCGTCTGCTCCACGCTG 4306
 ||||| ||||| ||||| ||||| ||||| |||||
 2186leasplAarGluLeuInolymethrAlaProLeuProThra 2200
 ||||| ||||| ||||| ||||| ||||| |||||
 4307 CCGCGCGCAAGCGCGCAACTGGAAGCGCAACAGCGCGGCGATC 4353
 ||||| ||||| ||||| ||||| ||||| |||||
 2200 snserPheLys.....GlnSerLysThrSerPheaspAlaGlyVal 2213

seq.name: SwissProt_40:YDEK_ECOLI

seq_documentation_block:

ID YDEK_ECOLI STANDARD; PRT; 1325 AA.

AC P32051; P76140; P77168;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical lipoprotein ydek precursor (ORF7).
 GN YDEK OR ORF7 OR B1510.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / M61655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Nakano K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saio N.,
 RA Sampei G., Seki Y., Sivasubram S., Yamaguchi H., Yakeda J.,
 RA Takemori G., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN [3]
 RP SEQUENCE OF 595-1325 FROM N.A.
 RX MEDLINE=94100243; PubMed=8274505;
 RA Cartwright P.J., Timms M.W., Lithgow T., Hoef P.B., Hoogenraad N.J.;
 RT "An Escherichia coli gene showing a potential ancestral relationship
 to the genes for the mitochondrial import site proteins ISP42 and
 MOM38.";

RL Blochim. Biophys. Acta 1153:345-347(1993).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 CC (potential).
 CC -1- SIMILARITY: TO E. COLI YDEK.
 CC -1- SIMILARITY: SOME, TO FUNGAL MITOCHONDRIAL IMPORT SITE PROTEINS
 CC ISP42 AND MOM38.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 CC FRAMESHIFT IN POSITION 653.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000248; AAC74583.1; -.
 CC EMBL: D90793; BAA15190.1; ALT_INIT.
 CC EMBL: D90794; BAA15197.1; ALT_INIT.
 CC EMBL: X73295; CAA51730.1; ALT_FRAME.
 CC PIR: S34315; S34315.
 CC Ecogene: EG11780; ydek.
 CC DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 CC DR Hypothetical protein; Membrane; Lipoprotein; Signal;
 CC KW Complete proteome.
 CC FT SIGNAL 1 18 POTENTIAL.
 CC FT CHAIN 19 1325 HYPOTHETICAL LIPOPROTEIN YDEK.
 CC FT LIPID 19 19 N-ACYL DIGLYCERIDE (POTENTIAL).
 CC FT CONFLICT 884 884 M -> K (IN REF. 3).
 CC FT CONFLICT 1317 1317 M -> S (IN REF. 3).
 CC FT SEQUENCE 1325 AA; 136514 MW; 26A3A06FA19AD7D CRC64;

alignment_scores:

Quality: 240.50 Length: 1492

Ratio: 0.362 Gaps: 80

Percent Similarity: 44.504 Percent Identity: 20.040

alignment_block:

us-09-303-518d-649 x YDEK_ECOLI ..

Align seg 1/1 to: YDEK_ECOLI from: 1 to: 1325

619 GATGAGATGAGCCCAATTAACCGCAAGTTCATAT...CATATGCAAG 665
 |||||..... ||||| ||||| |||||
 67 AspsnaspGlnIleThrAsnIleAspThrAspAlaIleTyrAspAlaTyr 83
 ||||| ||||| ||||| ||||| ||||| |||||
 666 TGGCATTCCTGG.....CTCGTTGGTGGA 691
 ||||| ||||| ||||| ||||| ||||| |||||
 83 rleuValGlyTrrPyrGlyThrGlyValLeuAsnIleLeuAlaGlyGly 100
 ||||| ||||| ||||| ||||| ||||| |||||
 692 AT.....ACCTTGGCAAAATGATCA 714
 ||||| ||||| ||||| ||||| ||||| |||||
 100 snAlaSerLeuThrThrIleThrThrSerValIleGlyAlaAsnGluAsp 116
 ||||| ||||| ||||| ||||| ||||| |||||
 715 GGTGGTGGCAGACCACTTA.....GGTAGAAATAATTAACATAG 758
 ||||| ||||| ||||| ||||| ||||| |||||
 117 SerGluGlyThrValAsnValLeuGlyGlyThrTrpPyrLeuTyrAsp 133
 ||||| ||||| ||||| ||||| ||||| |||||
 133 rGlysnAsnAlaArgProLeu.....AsnValGlyGlnSerIleTyrG 148
 ||||| ||||| ||||| ||||| ||||| |||||
 759 CCATATGATGTTTTTACCACAGAGGCTCATTTGGCGACAGTGGCTCAC 808
 ||||| ||||| ||||| ||||| ||||| |||||
 133 rGlysnAsnAlaArgProLeu.....AsnValGlyGlnSerIleTyrG 148
 ||||| ||||| ||||| ||||| ||||| |||||
 809 CATGTATTATCATATATGCCCAAAAGAGTGGTAAATTAATGAGGTA 858
 ||||| ||||| ||||| ||||| ||||| |||||
 148 lYThrLeuAsnIle.....LysGlnLysGlyHisValAspGly... 160
 ||||| ||||| ||||| ||||| ||||| |||||
 859 TGGCAACGCGCAACCCCTATATGAGAAAGACATGCGTTCGAGTGGT 908
 ||||| ||||| ||||| ||||| ||||| |||||
 161GlyTyrLeuArgLeuLysSerThrGly..... 170
 ||||| ||||| ||||| ||||| ||||| |||||
 909 TCGTAAGATGTTGTTATGATGAATCAATCTTTCGCGAGATGCCATTCAG 958

170 170
959 TATTCAGAACCGTCAAAATGGAATCTTTTAAACGAGATAT 1008
170 170
1009 AATGCGACGAGAAATCATGCAACATGACACATCTCTGCTAA 1058
171 ...GlyValGlyThrValAsnValGlyGlnAspSerValLeu... 184
1059 TAGATTAATAACACGACCGTTCAATGTTAATGTT... 1095
185ThrThrGluLeuPheGluIleGlySerTyrGly 196
1096TCTTATCCGAGACAGACGAAACGCTTTTCACTGTCAGCT 1140
196 hrGlySerLeuAsnIleThrAspGlyGlyValThrSerIleVal 212
1141 GGTGTCAACACTATGACACCCAGCTGAATATGAGAAATATTTCTT 1190
213 AlalleuGlyTyrGlnAla.....GlySerAsnGlyGlnVal 225
1191 TATTGACGAGAAAGGCAATTGATCTTACACACATC..... 1233
225 lValValGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 242
1234AATCAAGTGTGAGAGATTAATTTCCAAAGA 1266
242 leGluPheGlnIleGlyAsnGlnGlyThrGly..... 252
1267 GATTTACGCTCGCCTGAAATTAACGAACCTGCGCAAGCGCGGCGT 1316
253GluAlaThrIleArgGlyGlyGlyGlyGlyGly 261
1317 TCATATCAGTGAAGACAGTACCTTGTGAAGATTAACGCGGCGCA 1366
261 uValThrAlaGlyAsnThrIleIleGlyGlyAsnAlaThrGlyIle... 276
1367 ACAGCCGCTGTCCAAATCGGCAAGGACGCTGACGTTCAACCAAA 1416
277GlyThrLeuAsnValGln..... 282
1417 GGGGAAACCAAGGCTCGATCAGCGT..... 1443
283 ...AspGlnAspSerValIleThrValArgArgLeuTyrAsnGlyTyrPhe 298
1444 GCGCAGCTACAGTCATTTTGGATCAGCAGCAGCAGATTAAGCAAAA 1492
298 eGlyAsnGlyThrValAsnIleSerAsnAsnGlyLeuIleAsnAsnGly 315
1493 AACAAAGCCTTAGTGAATCGGCTGTGTCAGCGC...AGGGTACGCTG 1539
315 lU.....TyrSerLeuValGlyValGlnAspGlySerHisGlyValVal 329
1540 CAAGTGAATCCGATATCAGTTCAACCCGCAAACTATTTCCGGCTT 1599
330 AsnValThr.....AspGlySerGlyHisThrPasnPh 339
1599 TCGCGGCGGAGCTTGGATTTAAAGCGGCACTTTCCTTCACCGCTA 1639
339 eLeuGly.....ThrGlyGlnAlaPheArgTyrIleTyrI 351
1640 TTCAAAATACGATGAAGGCGGATGTGCAACCCACATCAAGCAAAA 1689
1690 GAATCCACCGTTACCATTTACAGCAATTAAGATATGCTCAACCGCAAA 1739
368 AspSerGlyIleIleThrAlaGlyMetGlyThrGly...ThrGlyAs 383
1740 TTAACAACGCTTGATAGCAAAAAGAAATTCCTACACGCTTGGTTG 1789
383 nIleThrValIysAsp..... 388
1790 GCGAAGAAATATACGACCAAAACGAGCGGCGCTCAACTGTTTACAG 1839
389 ...LysAsnSerValIleThrAsnLeuGlyThrAsnLeuGlyTyrAsp 403
1840 CCGCGCGCAGAA.....GACCGACCCGCTGCTTCC...GG 1874
404 GlyHisGlyGluMetAsnIleSerAsnGlnGlyLeuValIleSerAsn 420
1875 CGGACCAATTTAAAC.....GGCAACATCAGCAAAA 1906
420 yGlySerSerLeuGlyTyrGlyGluThrGlyValGlyAsnValSerIle 437
1907 CAACGCGCAACTGTTTTCAGGCGGACACACCGCAGCTTCAAT 1956
437 hrThrGlyGlyMet..... 441
1957 CATTTAAACGACATTGGTCGCAAAAAGGCGCATTCCTCGCGGGAAT 2006
441 441
2007 CGTGGGACACGACCTGATCAACCGCACATTTAAAGCGAA..... 2049
442 ...TrrpGlu.....ValAsnLysAsnValTyrThrThrIleGly 454
2050AATTCGCAATTTAAAGCGGAGCGGCGGCTGTTCC 2085
454 alAlaGlyValGlyAsnLeuAsnIleSerAspGlyGlyIleValSer 470
2086 CGCAATGTGCGCAAGTGAAAGCGGATTTGGCATTTGACCAATCAGCCCA 2135
471 GlnAsnIleThrPheLeu...GlyAsp..... 478
2136 AGCAGTTTTGGTGTGCGACCGCATCAAGCCACCAACATCTGTACAGT 2185
479LysA 480
2186 CGGACTGACGAGGTGTGCAAAATGTGTGCAAAAACCATTAACGACAT 2235
480 lAserGlyIleGlyThrLeuAsnLeuMetAspAlaThrSerPheAsp 496
2236 AAGTATGTGCTTCAATGACTGAAGACCGACATCAGCGCATGTGCATCT 2285
497 ThrValGlyIleAsnValGlyAsnPheGly...SerGlyIleValAsnVal 512
2286 TGCGGATCAGCTCATTTAAATCTCAGAGGCTTGCCACACTCAAGCGCA 2335
512 lSerAsnGlyAlaThrLeuAsnSerThrGlyTyrGlyPheIleGlyIle 529
2336 ATCTTAGTGCACAAATGCG.....GATACAGCTTAT... 2364
529 snAlaSerGlyLysGlyIleValAsnIleSerThrAspSerLeuThrPasn 545
2365ACAGTCAGCCACACGCGC..... 2382
546 LeuLysThrSerSerThrAsnAlaGlnLeuLeuGlnValGlyValLeuG 562
2383 .ACCAAAAACGCAACTTACCTCTGCGCAATGCCAAGCAACATTTA 2431
562 yThrGlyGlyLeuAsnIleThrThrGlyGlyIleValIleAlaArgAsp 579
2432 ATCAAGCACCATTTAAAC.....GGCAACACATGCGGCTTGGCG 2469
579 hrGlnIleAlaLeuAsnAspLysSerIleValAspValArgValAspGly 595
2470 AATGCTTCA.....TTTATCTAAGGACACCGCTTACAAAA 2507
596 GlnAsnSerLeuLeuGluThrPheAsnMetTyrValGlyThrSerGly 612
2508 CGGAGTGTGACGCTTTCGCGCAACGCTAAGCAACGTA..... 2547
612 rGlyThrLeuThrLeuThrAsnAsnGlyThrLeuAsnValGlyGlyG 629

```

2547 ..... 2547
629 luValIyrlEuGlYAlPheGlUPrOAlaValGlYThLeuAsnIleGly 645
2548 ...AGGCAT.....TCGCACTCAACGGTAATGTCTCCAGCCGA 2585
646 AlAlAlHISglYlunAlAlAlAlAspAlaglyPheIleThrAsnAlaTh 662
2586 TAAGCAGATATTC.....CATTTTG 2605
662 rlySValIglUheglYleuclYglUglYAlPheValPheAsnHISThrA 679
2606 AAGCGAGC.....CGCTTACCGGACAAATC 2631
679 snAsnSerAspAlaGlYlYrGlInValAspMetleuIleThrcIlyAspSp 695
2632 AGCGGCGGCAAGATACGGCATTACCTTAAAGACAGCGAATGGACGT 2681
696 lysAspIlyLys.....ValIleHIS.....AspAlaGlYHISThrVa 708
2682 G.....CGTCAGGCGACGAATTAGCAATTTAACTTGACAAAGCCA 2725
708 lPheAsnAlaGlYAsnThrYlSerGlYStHrleuValAsnAspGlYL 725
2726 CCATTACACTCAATTCGCTATCGCAGATCGGCGAGGCGCAAC 2775
725 euleuThrIleAlaser.....HISThrAlaSpGlYAl...Thr 737
2776 GCGAGTCGCAAGATGGCGCGCGCGCTTCGCGCGCTTCGCGCGCTC 2825
738 GlyMetGlySer.....Se 742
2826 CCTATTATCCGTACACCGCCAACTTGAGATCCGTTCAACAG. 2874
742 rGlUValIhrlleAlAsnProglYThrleuAspIleLeuAlaSerThra 759
2875 .....CTGACGGTAACGGCAATTGACGGTGACGGTCA 2910
759 snSerAlaGlYAspYlThrleuThrAsnAlaLeuylSglYAspGlYleu 775
2911 TTCGCTTTATG.....TCGAACTCTTCGGCTACCGCCAGCA 2948
776 MetAlaGlYAlGlnleuSerSerAspIlySmetPheglYpHe..... 789
2949 CAATTTGAAGCTGGCGGAAAGTCCGAAGACATTACACTTGGCGGTCA 2998
790 .....Thr 791
2999 ACAATACCGGCAACGACCTGCAAGCTCGAACAATTGACGGTAgTGA 3048
791 lAlAlaThrcIlyThrcIlyPheAlaGlYAlAlaGlInleu..... 803
3049 GGAAGAAAGACAAACCGCTGTCCGAAAACCTTAATTTCACCCGCAAAA 3098
804 ...LysAspSer.....ThrPheThrleuGlUAr 812
3099 CGAACACGTCGATGGCGGCGGTGCGTTACCACTCATCCGCAAGACG 3148
812 gAsp.....AsnThrAlaIleuThrHISAlaMetleuGlInSerAspS 827
3149 GCGAGTTCGCGCTCATTAATCCGGTCMAAGCAAGACCTTCCGCAAAA 3198
827 erclUAsnThrTrSerValIySValIglUglInSerIleGlYglYleu 843
3199 .....CTCGCAAGCGCAAGCAACCAAAACAGC 3227
844 AlAlaMetAsnGlYglYThrIleIlePheAspTrAspIlePrOAlAlaIaTh 860
3228 GGAAGAAAGACGCGCAAGACCTTGACCGGCTGATTGGCGCGCGCG 3277
860 rLeuAlaIglUglYlYrIleSerValAspThrleuValIglYAlaGlY 877
3278 ATCCGCTGGAAGACAGCAAGC..... 3300
877 sPlylThrTrPlySglYArYAsnTYrGlInValAsnGlYThGlYAspVal 893
3301 .....GTTCGCCAACCGCGCCGCGAGCGGAGGGAAT..... 3336
894 leuIleAspValProlySPrOTrPAsnAspPROMetAlaAsnAspProle 910
3337 .....GTTCGCAT..... 3345
910 uThrThrleuAsnleuGlInHISAspSerHISValIglYAlGln 927
3345 ..... 3345
927 euValIySAlaGlInThrValIleGlySerGlYSerleuThrleuArg 943
3346 ...ATCGAGCGGAGAGAGAAAAACCGGTGACGGGATTAAGACAC 3392
944 AspLeuGlInGlYAspIle.....ValGlUAlaAspIlySThrle 956
3393 CGGCTTGGCGGAACAGCGCGAAGCGAAACCGG.....C 3427
956 uHISlIleAlaGlInAsnGlYThrValValAlaIaGlUglYAspYrGlYphea 973
3428 CGGCTACACCGGCTTCGCCCGCGCGCGCGCGCGGATTTGGCG 3477
973 rglUeUThrThrAla..... 977
3478 CAACGTGAACCGCAACCGGACGCCCAACCGGACCGGACTGATGACCG 3527
978 .....ProglYAsnGlYleu..... 982
3528 TTATGCCATAGCGGTTTG..... 3546
983 .TYValAsnTYrGlYleuYlSAlaLeuAsnIleHISglYglInLysL 999
3547 .....AGTGAATTTTCGCCGACGCAACAGCGT 3576
999 euThrleuAlaGlInHISglYAlaTYrGlYAlaThrlaAspMetSer 1015
3577 TTCGCCGTACAGAGCAATTAGACCGCGTATTTGCCGAAGACCGCGCA 3626
1016 AlAlylSlleGlYglYglUglYAspLeuAlAlaAsnThrValArGlInVa 1032
3627 CGCGGTTTGCAACAGCGGATCCGGGACACCAACACTTACCTTCCAG 3676
1032 lSerleu...SerAsnGlYglInAsnAsp.....TYrGlInGlY... 1043
3677 ATTCCGCGCTACCGCCCAACAAACGACGCTGC..... 3711
1044 .....AlaThrTYrValGlInMetGlYThrleuArgThrAspAlaAspGlY 1058
3712 CAATCGGTATGCAAGAAAACCTCGGACGCGCGCGCTCGGCATC..... 3756
1059 AlAlaGlYAsnThrArGlInleuAsnIleSerAsnAlAlaIleValas 1075
3757 .CTGTTTTCGCAACCGGACCGGAAACACTTCGACAGCGGCAATCGCA 3805
1075 pleuAsnGlYSerThrGlInThrValGlUThrPheThrGlYglInMetGlYs 1092
3806 ACTCGGACAGGCTTCCGAGCGCGCGGCTTTTCGGCAATACGATCGAC 3855
1092 erThrValleuPheylSglUglYAlaIleuThrValAslYsGlY..... 1106
3856 AGGTTCTACATCGGATCAGCGCGGCGGCTTTTACAGCGGACGCT 3905
1107 .....GlyIleSerGlInGlYlUleuThrGlYglYAsnle 1119
3906 TTCAGACGCGATCGAGGCAAAATCCGCGCGCGCTGCTGATTACGCA 3955
1119 u...AsnValThrGlYglYThrleuAlAlleU.....GlyL 1131
3956 TTCAGGACAGATACGCGCGGCTTTCGGCGGATTCGGATGGAACCGCAC 4005

```


FT VARIANT 815 815 A -> V (IN STRAIN ML 308-225).
FT VARIANT 824 824 C -> S (IN STRAIN ML 308-225).
FT VARIANT 829 835 LNVHTS -> MNLVNA (IN STRAIN ML 308-225).
FT VARIANT 845 847 OGCT -> LGA (IN STRAIN ML 308-225).
FT VARIANT 855 855 S -> T (IN STRAIN ML 308-225).
FT VARIANT 888 888 Q -> L (IN STRAIN ML 308-225).
FT VARIANT 1025 1025 S -> I (IN STRAIN ML 308-225).
FT CONFLICT 61 63 ETV -> TTT (IN REF. 5).
SQ SEQUENCE 1039 AA; 106841 MW; 5170D647C8DEBBD CRC64;

alignment_scores:
Quality: 233.50 Length: 1317
Ratio: 0.408 Gaps: 64
Percent Similarity: 43.508 Percent Identity: 20.046

alignment_block:
US-09-303-518D-649 x AG43_ECOLI ..

Align seg 1/1 to: AG43_ECOLI from: 1 to: 1039

682 GTTCGTGGCAATACCTTGGCAAAATGGA..... 711
63 Valasnlglythrleualasnhisaspasnglnllevalphee1yth 79
712 .TCAGGTGTGGCAGCTCACTAGGTAGTGAATAAATTAACATAGCC 760
79 rthrasnlglymethrileserthrgly.....Leuglutrylglyp 93
761 CATATGTTTTTTTACCAAGAGAGGCTCATTTGGCGACAGTGGCTCACA 810
93 roaspasnglnualasnthrlelgly..... 101
811 ATGTTATCTATGATGCCCAAAAGCAAAAGTGTATTAATGAGCTATT 860
102Glntrvalglnasp1y..... 107
861 GCMAAGCGGCAACCCCTATATAGAAAAGCAATGGCTTCCACGTGGTTC 910
108 .g1ythrilaasnlystrhrthval1thrserglyleuglnhrvala 124
911 GTAAGATTGTTCTATGATGAA.....ATCTTGGTGGAGATACCCAT 954
124 snp1rog1glyservalserasphrval1leaser1ag1glygln 140
955 TCAGTATTCCTACGAACCGCTCAAAATGGAAATACCTTTTAAACGACA 1004
141 Serleu.....Gln1y1arq...Alavalaasnthrth 150
1005 TAAATATGGC.....ACAGGA 1021
150 rleuasnlglylglnlntprmeth1sg1ug1yAla1llealthr1g1yT 167
1022 AA...ATCAATGCCAAGCATGACAAATCTCTGCTTAATAGATTAAA 1068
167 hrval1leasnsp1ystrg1ytrglnval1val1ys1rog1yThVala1a 183
1069 ACACGACCGCTCAATGTTTAAATGTTCTTATCCGAGACAGACAAGA 1118
184 Thr1asph1rval1.....Valasnthr1g1y..... 191
1119 ACCTGTTTATCATGCTCAGGTGTGTCAACAGTTATGACCCACAGTGA 1168
192Alag1ug1y1g1y.....Proasp1ag 199
1169 ATATGGAGAAATATTTCTTTATGACGAAGAAAGCGAATGATA 1218
199 l1uasnlg1y1asph1r1g1y1n1pheval1.....Arg1y1asph1a1val 212
1219 CTTACGACGACATCAATCAAGT.....GCTGAGAGA.. 1251
213 Arg1thr1leasn1y1asnlg1y1arq1n1lleval1r1gal1ag1ug1y1th 229

1252TTATATTCAGAGAGATTTTACGCTCGCTG 1285
229 r1alaasn1r1thr1val1y1r1al1ag1y1a1ysp1n1thr1val1h1s1g1y 246
1286 AAATATACGAACCTTGGCAAGCGCGGCG.....GTTCAAT..... 1320
246 l1ala1leuasn1r1thr1leuasn1y1g1y1n1y1val1h1s1asn1y 262
1321 ...ATCAGTGAAGACATGACGCTTACT.....TGGAAAGTA..... 1353
263 g1y1thr1ala1ser1asph1r1val1val1asn1ser1asph1y1n1lleval1y 279
1354 .AACGGCGTGGCAAAACGACCGCTGTCCAAAATGCGCAAGACGCTGC 1402
279 s1asn1y1g1y1al1a1ag1y1asn1r1thr1Val1asn1n1ys1g1y1r1eug 296
1403 ACGTTCAGCCCAAAAGGGAACCAAGCTGCATGAGCGTGGCGACGCT 1452
296 l1val1asph1a1ag1y1y1thr1al1a1thr1asn1val1Thr1leu1ys1g1n1y1y 312
1453 ACAGTCAATTTTGATGATCAGCAGCAGCATTAAGCCAAACAAAGCCTT 1502
313 Alaleuval1thr1ser1thr1al1a1a1thr1val1thr1g1y..... 325
1503 TAAGTGAATCGGCTTGGTGTACGCGGAGGCTACGCTGCAACGATGCGC 1552
325 easn1arq1y1al1a1p1heser.....Val1al1g1ug1y1ys1al1a 339
1553 ATATATGCTTCAACCCGCAACACTATTTTGGCTTGGCGAGAGCT 1602
339 sp1asn1val1al1leu1g1uasn.....Gly1y1arq 348
1603 TTGGAT...TTAAACGGGCAATCGCTTGGTTCACCGTATTCAAATATC 1649
349 Leu1s1p1al1leu1thr1g1y1n1st1hr1al1a1thr1asn1r1arq1val1..... 362
1650 CGATGAAGGGCGCGATGTTGTCAACCAACATCAAGCAAAAGATCCAGCG 1699
363 .Asp1asp1y1g1y1thr1leu1s1p1al1arq1asn1y1g1y1thr1al1a1thr1 379
1700 TTACCATTTACAGGCATTAAGATATTGCTACAAACCGCAATTAACACAGC 1749
379 al1ser1met.....Gly1asn1y1g1y1al 386
1750 TTGGATACCAAAAAGAAATGCTTACACAGGTTGGTGGCGAAGAA 1799
387 Leu1leu1al1a1asp1ser1g1y1al1a1a1val1ser1y1..... 397
1800 TACGACCAAAACGAACGGCGGCTCAACCTTGTATTACACCGCGCGCAG 1849
398 ...Th1arq1ser1asph1y1s1al1a1p1heser1le1g1y1g1y1n1la1a 413
1850 AAGACCGGACCCCTGCTTTCGCGGACCAAT.....TAAACGGC 1893
413 sp.....Alaleu1met1leu1n1l1u1s1y1ser1p1het1r1leu1asn1a1a 427
1894 AACATCAACGCAACA.....AACGCAACCTGTTTTCACGG 1931
428 Gly1asph1r1al1a1thr1asph1r1Thr1Val1asn1y1g1y1leu1p1het1r1al1a.. 443
1932 CAGACCAACACCGCAGCGCTCAATCATTTAAACGACCATTTGGTCGCAA 1981
443 443
1982 AAGAGGCAATTCCTCGCGGGAATCGTGGGACAAAGACTGATCAAC 2031
443 443
2032 CGCACATTTAAAGCGAAATTCCAATTAAGCGGACAGCGGTGCT 2081
444Arg1y1y1y1..... 446

```

2082 TTCGCCAATGTTGCCAAAGTGAAGCGATTGGCATTTGACCAATGACG 2131
447 .....ThrlleuAlaGlyThrThrlleuSnsnIYA 458
2132 CCCAAGCAATTTTGTGTGCGACCGCATCAAGCCACACAATCTGTACA 2181
458 Ia.....IleLeuThr 461
2182 CGTTGCGACTGACGGGTCTGACAAATTTGTGCAAAAAACCATTTACCGA 2231
462 LeuSerGly.....LysThrValAsnAs 469
2232 CGATTAAGTATGCTTCATTACTAAGACCGACATCAGCGGCAATCTCG 2281
469 nAsp.....ThrLeuThrIleArgIleGlyIAspAlaI 480
2282 ATCTTGCGGATCAGGCTCATTTAAATCTCACAGGCTTGCACACTGAC 2331
480 euleuGlnGly.....GlySerLeuThrGlyAsn 489
2332 GGCATCTTATAGTGCMAATGGCGATACAGCTTATACAGTGCACCAACGC 2381
490 GlySerValGluYsserGlySerGlyThrLeuThrValSer...AsnTh 505
2382 CACCCAAAGCGGCACTTAGCCTCGTGGGCAATGCCCAAGCAACATTTA 2431
505 rThrLeuThrGlnIysAlaValAsnLeuSnsnIglYThrLeuThrLeuA 522
2432 ATCAAGCCACATTTAAAGCGCACACATCGGCTTCGGGCAATGCTTCAT 2481
522 snAspSerThrValThrThrAspValIleAlaGlnIArgIlyThrAlaLeu 538
2482 AATCTAAGCGACCAAGCGCTACAAAAGCGCACTG..... 2517
539 LysLeuThrGlySerThrValLeuSnsnIglYAlaIleAspProThrAsnVa 555
2518 .ACGCTTTCGGCAAGCTTAAGCAAGCAATGCCATTCGCACTCAACG 2566
555 lThrleuAlaSerGlyAlaIleThrTrpAsnIleProAspAsnAlaThrValG 572
2567 GTATATGTC.....TCCCTAAGCCGATTAAGCAATGCTTCATTTGAAGC 2610
572 lnservalValAspAspLeuSerHisAlaGlyGlnIleHisThrSer 588
2611 AGCCGCTTAAACGGNCAATCAAGCGCGCAAGATACGGCATTCACATT 2660
589 ThrArg...ThrGlyLysPhe..... 594
2661 AAAAGACAGCGAATGACGCTGCGCTCAGCGACGGAATTTAGCAATTTAA 2710
595 .....ValProAlaThrLeuIysValIysAsnLeuA 605
2711 ACCTTGACAAGCGCAACATTTACATCAATTCGCGCTATCGCCAGATGCG 2760
605 snIglYlmsnIglYThrIleSerLeuArg..... 614
2761 GCAGGGGGCAAAACCGGACGTGCGACAGATGCGCGCGCGCTTGCGG 2810
614 ..... 614
2811 CGGTGCGCGCTTCCCTATTATCCGTTACACCGCAACTTCGTTAGAT 2860
615 .....ValArgProAspMetAlaIlnA 622
2861 CCCGTTTCAACAGCTGACGTTAAACGGCAATTTGAACGCTGACGGACA 2910
622 snAsnAlaAspArgLeuValIleAspIly.....GlyArgAlaThr 635
2911 TTCGCTTATGTCGAACCTCTTGCTTACCGACGACCAAAATTTGAAGT 2960
636 GlyLysThrIleLeuAsnLeuValAsnAlaGlyAsnSerAlaSerGlyLe 652
2961 GCGGAAAGTTCCGAGGCACTTACACTTG...GCGGTCAACAATACCG 3007

```

```

652 uAlaThrSerGlyLysGlyIleGlnValValGlnAlaIleAsnIlyAlaT 669
3008 GCAAGCAACCTGCAAGCCTCGAACAAATTTGACGGTAGTGGAAGAAAAAGC 3057
669 hThrGluGluIlyAlaPhe..... 675
3058 AACAAACCGTGTCCGAAAACCTTATTTCAACCTGCAAAACGACACAGT 3107
676 .....ValIleIlyAsnArgIle 681
3108 CGATGCGCGCGCTGCGGCTTACCAACTC..... 3135
681 uGlnAlaGlyIlyAlaPheAsnIlySerLeuAsnArgAspSerAspIleT 698
3136 .....ATCCGCAAGACGGGAGGAGTTCGCTGCATTAATCCGTCAAAGAA 3180
698 rPlyrLeuArgSerGluAsnAlaIlyrArgAlaIlyAlaIProleuTyAla 714
3181 CAAAGGCTTCCGCAAAACTCGGCAAGCGCAAGCCAAAAACAGCGCGA 3230
715 SerMetLeuThr..... 718
3231 AAAAGACAGCGCGCAAGCCTTAGCGCGCTGATTTGCGCGCGCGGATG 3280
719 .....GlnAlaMetAspPlyrArgIlyAlaIlyAlaIlySerArg... 731
3281 CCGTCGAAAGACAGAAAGGTTGCCGAACCGCGCGCGAGGACAGC... 3327
732 .....SerHisGlnThrIlyAl 737
3328 ..GGGAAATGTCCGCAATTATGACAGCGGAGAAAGAAAAACGGGT 3374
738 AsnGlylmsAsnAsnSerVal...ArgLeuSerIleGlnIlyGlnIle 753
3375 GCAAGGCGATTAAGACACCGCCTTGGCGAAACAGCGCGCAAGCGAAACC 3424
753 uGlnIlsAspAsnAsnIlyIleIleIleIleIleIleIleIleIleIle 769
3425 GCGCGGCTACACCGCCTTCCCGCGCGCGCGCGCGCGCGGATTTG 3474
770 ..SerIlySerIlyrGlyPheValArgLeu..... 778
3475 CCGCAACTGCACACCCCAACCGGACCGCCACCGCGCACTGATCAG 3524
779 .....GluIlyAspLeuMet... 783
3525 CCGTATGCGCAATACGCTTTGAGT.....GAATTTTCGCGCA 3562
784 .ArgThrGluValAlaIlyMetSerValThrAlaGlyAlaIlyrGlyAla 800
3563 GCGTCAACAGCGCTTTCGCTACAGGACGATTAAGCCCGGATTTGCC 3612
800 lAGlYHisSerSerValAspValIlyAspAspArgIlySerArgAlaGly 816
3613 GAAGCGCGCGCAACCGCTTGGACAAGCGGACATCCGGGCGCAACCA 3662
817 ThrValArgAspAspAlaGlyCysLeuGlyIlyrIleuSnsnLeuValH 833
3663 CTACCGTTG.....CAAG 3676
833 sThrSerSerIlyLeuThrAlaAspIleValAlaIlyGlnIlyThrArgHis 850
3677 ATTTCCGCGCTTACCGCAACCAACGACCTGCGCCAAATCGGTATGACG 3726
850 erMetIlyAlaIleSerSerAspAsnAsnAspPheArgAlaIlyrGlyrGly 866
3727 AAAAAGCTCGGACGCGCGCGCTGCGCATCTTTTCGACCAACCGGAC 3776
867 Trp...LeuGlySerLeuGluThrGlyLeuProPheSer.....IleH 880
3777 CGAAAC.....A 3784

```

```

RT surface layer protein of Rickettsia typhi..":
RL Gene 133:129-133(1993).
RN [2]
RP PARTIAL SEQUENCE.
RC STRAIN=WILMINGTON;
RX MEDLINE=92114896; PubMed=1370573;
RA Ching W.M., Carl M.W., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNBR fragments of
the S-layer protein antigens of Rickettsia typhi and Rickettsia
provakexii.";
RL Immunol. 29:95-105(1992).
RN [3]
RI IDENTIFICATION OF CLEAVAGE SITE.
RP MEDLINE=92104668; PubMed=1729180;
RA Heckstadt T., Messer R., Cleplak W., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kDa outer
membrane protein of rickettsiae: identification of an avirulent
mutant deficient in processing."
RL Infect. Immun. 60:159-165(1992).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKEITSIAL
VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
S-LAYER WITH HEXAGONAL SYMMETRY.
CC -1- SIMILARITY: BELONGS TO THE RICKEITSIAE OMPA/OMPB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
its European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sdb.ch/announce/
or send an email to licenseds@sib.ch).
-----
DR EMBL; L04661; AAA84987.1; .
DD InterPro; IPR003858; rompa_rompb.
DR Pfam; PF02708; rompa_rompb; 1.
KW Antigen; S-layer; Transmembrane; Cell wall.
FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
FT TRANSMEM 1415 1429 32 KDA BETA PEPTIDE.
FT CONFLICT 657 657 MEMBRANE ANCHOR (POTENTIAL).
FT CONFLICT 842 842 V -> N (IN REF. 2).
FT CONFLICT 1071 1071 G -> A (IN REF. 2).
FT CONFLICT 1306 1306 G -> S (IN REF. 2).
SQ SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;

alignment_scores:
Quality: 233.00 Length: 1454
Ratio: 0.350 Gaps: 76
Percent Similarity: 45.736 Percent Identity: 20.495

alignment_block:
US-09-303-51BD-649 x OMPB_RICTY ..

Align seg 1/1 to: OMPB_RICTY from: 1 to: 1645

661 GCAGTGCGCATTTCTGGCTGCTGTGGGCACAAATCACTTCACAAAATG 710
||||| | :||| ||||| | | | | |
251 AlasahAlaleuasnleuglnelylgelivashnthrllesnpheasnsl 267
711 A...TCAAGTGCGCACACTCAACTAGTAGTAAGAATAATTAAACATA 757
| | :||| | | | | | | | | | | | |
267 yagaspelgylthrglylysleuallevalserlyasnsnglyasnl 284
758 GCCCATATGGTTTTTACCACAGAGAGCTCATTTGGCGCAGATGGCTA 807
:: :: :: ||||| ::| | ::| ::|
284 hrglupheasn.....ValnhtglyseuleuglyglasluLeuys 297
808 CCATGTTTATCTATCAT.....GCCCAAAGCAAAAGTGTAATTAA 851
::::: ::||| | | | ::| |
```


[illegible]

1569 CGACAAACTCTATTTCCGCTTTCCGCGCGGACAGCTTTGGATTAA..... 1611

564AlaIleHisPrpGlnAlaAsnGlyIleThrIleGlnLeuThrIleTser 579

1612AACGGGCAATCGGTTTCGTTCCACCGCTTTCAAAATCCAGTGA 1656

579 htrGlnAsnAsnIleLeuValAspPheAspLeuAspValThrThrAspGln 595

1657 GGGCGGATGATTGTCACAAACAATCAAGACAAAGAAATCCACCGTTACCAT 1706

536 ThrGlyValValAspAlaSerSerLeuThrAsnAsnGlnThrLeuThrIle 612

1707 TACAGCGAATTAAGATTTGCTCTCAACCGCGGCATTAACAACAGCTTGATTA 1756

612 eAsnGlySer.....IleGlyThrIleGlyAlaAsn..... 622

1757 GCAAAAAGAAATTCCTTACACGCGTTGGTTGGGAGAAAGATACGACC 1806

623 1807

1807 AAAACGAAGCGGCGGCTCAACGCTTTGTTACAGCCCGCGGAGAAAGACG 1856

624 LysThrLeuGlyArgPheAsnVal.....GlySerSerIle 635

1857 CACCGCTGCTCTTCCGCGGGAACAATTTAAACGGCAACATCCAGCGAA 1906

635 sThrIleLeuAsnAlaGlyAspValAlaIleAsnIleuValuIleGluVal 652

1907 CAACGGCAACGCTTTTTCACGCGGACGACCAACGCGACCGCTTAC... 1953

652 sAspArgLysSerValHisLeu.....ThrHisAsnThrTyrLeu 664

1954AATCATTTAAACGACCATTTGGTGGCAAAAGAGG 1988

665 IleThrLysThrIleAsnAlaHisn..... 673

1989 CATTCGTCGGGGGGAATGCTGTGGACAGACTGCATCAACCG...A 2035

674GlnGlyLysIleIleValAlaAlaAspProIleAsnThrAsp 688

2036 CATTTAAAGCGAAATCTCCAAATTTAAAGCGGACGCGGCGTGTTC 2085

688 htrAlaLeuAlaAspGlyThrAsnLeuGlySerAlaGluSerProIleSer 704

2086 CGC.....AATGTTGCCAAAGTGAAGCGAT.....TGCATTT 2120

705 AsnIleHisPheAlaThrLysAlaAlaAsnGlyAspSerIleLeuHisIle 721

2121 GAGCAATCAACGCGGCAAGCAGATTTTGGGTGGCGACCGCATCAAGGCACA 2170

721 eGlyLys.....GlyValAsnLeuTyrAlaAsnAsnI 732

2171 CAATCTGTACAGCTTCCGAC.....TGGACGGGTTCG 2202

732 LeThrThrThrAspAlaAsnValGlySerLeuHisPheArgSerGlyGly 748

2203 ACAAAATGTGTGCAAAAAACCATTAACGAGCAATTAAGTATTCCTTCAT 2252

749 ThrSerIleValSerGlyThrValGlyGlyGlnGlnGlyLeuLys.... 763

2253 GACTAAGACGCAATCAACGCGCAATGTGCATCTTGGCGATCAACGCTCAT 2302

764LeuAsnAsnLeuIleLeuAspAsnGlyThrThr 775

2303 TAAATCTCAGAGGCTTGCACACCTGACAGCGCAATTTAGTCAAAATGCG 2352

775 alyAspPheLeuGlyAspIleThrPheAsnGlyGlyThrLysIleGlnGly 791

2353 GATACAGCTTAATACAGTCAACGCGCAAC..... 2379

792 LysSerIleLeuGlnIleSerSerAsnTyrIleThrAspHisIleGlu 808

```

2380 .GCCACCAAAAGGCAACCTTAGCCTGGGCAATGCCCAACACAT 2428
      |||
      |||
      |||
808 rAlaAspAsnThrGlyThrLeuGluPheValAsnThrAspProIleThr. 824
      |||
2429 TTAAATCAGCCACATTAAAGGCAAC..... 2454
      |||
      |||
      |||
825 .....ValThrLeuAsnLysGlnGlyAlaTyrPheGlyValLeuLys 838
      |||
2455 .....ACATCGGCTTCGGGCAATGCTTCATTTAACTAAAGCAG.. 2493
      |||
      |||
      |||
839 GlnValMetValSerGlyProGlyAsnIleAlaPheAsnGluIleGlyAs 855
      |||
2494 .....CAGCCGCTA.....CAAAAGGCA 2512
      |||
      |||
      |||
855 nGlyValAlaHisAlaIleAlaValAspSerIleSerPheGluAsnAla 872
      |||
2513 GTCCTG.....ACGCTTTCGGC..... 2529
      |||
      |||
      |||
872 eLeuGlyAlaSerLeuPheLeuSerGlyThrProLeuAspValLeu 888
      |||
2530 AAGCTTAAGCAACGTAACGTAATCCGCACTCAACGTAATGTCCT 2579
      |||
      |||
      |||
889 ThrIleYSerThrValGlyAsnGlyThrValAsp...AsnPheAsnAl 904
      |||
2580 AGCCGATAGGCAGATATTCATTGTAAGACGCCCTTACCGGACAA 2629
      |||
      |||
      |||
904 AProIleLeuValValSerGlyLeuAspSerMetIleAsnAsnGlyGln 921
      |||
2630 TCACCGGGCGCAAGGATACGCAATACACTTA.....AAGACAGC 2670
      |||
      |||
      |||
921 aIleGlyAspGlnLysAsnIleIleAlaLeuSerLeuGlySerAspAsn 937
      |||
2671 GAATGAGCGCTCGCTACGACGCAATTTAGCAATTA.....AA 2711
      |||
      |||
      |||
938 SerIleThrValAsnSerAsnThrLeuTyrAlaGlyIleArgThrThrTy 954
      |||
2712 CCTTGACAAAGCCACCATTAACATTCGCTATCGCCAGCAT.... 2757
      |||
      |||
      |||
954 sThrAsnGlnGlyThrValThrLeuSerGlyGlyIleProAsnAsnPro 971
      |||
2758 .....GCGCAGGGGCGCAAAACGCGCAGTGCACAGATGGCGC 2796
      |||
      |||
      |||
971 LyThrIleTyrGlyLeuGlyLeuGluAsnGlyAspPro..... 983
      |||
2797 CGCCCGCTTCGGCGCTTCGGCGGTCCTTATTCGTTACACCGCC 2846
      |||
      |||
      |||
984 .....LysLeuLysGlnValThrPheThr 991
      |||
2847 AACTTCGTAGATCCGTTTCAACACGCTG..... 2877
      |||
      |||
      |||
991 rThrAsp.....TyrAsnAsnLeuGlySerIleIleAlaThrA 1004
      |||
2878 .....ACGCTAAC..... 2886
      |||
      |||
      |||
1004 sNValThrIleAsnAspAspValThrLeuThrThrGlyGlyIleAlaGly 1020
      |||
2887 .....GCAATTTGAACGTCAGGAGAC 2909
      |||
      |||
      |||
1021 ThrAspPheAspGlyLysIleThrLeuGlySerIleAsnGlyAsnAlaAs 1037
      |||
2910 ATTCCGCTTATGTCGGAACCTTTCGGCTACCGCAGCACAATTTGAAGC 2959
      |||
      |||
      |||
1037 nValLysPheValAspArgThrPheSerHisProThrSerMetIleValS 1054
      |||
2960 TGCGGGAAGTTCGGAAGCAGCTTACAC.....TTGCGCGTC 2997
      |||
      |||
      |||
1054 eThrLysAlaAsnGlnGlyThrValThrTyrLeuGlyLysnAlaLeuVal 1070
      |||
2998 AACAAATACCGCAAGAA.....CCTGCAAGCTTCGAACAATTAACGCT 3041
      |||
      |||
      |||
1071 GlyAsnIleGlySerSerAspIleProValAlaSerValArgPheThr.. 1086
      |||
3042 AGTGAAGAAAGAAACAAACAAACCGCTGTCGAAACCTTAATTTCACCC 3091
      |||
      |||
      |||
1087 .....GlyAsnAspSer.....GlyValGlyLeuGlnGlyAsnI 1098
      |||
      |||
      |||
3092 TGCAAAAGCAACACGTCATCGCGCGCGCTTACCAACTATCGCG 3141
      |||
      |||
      |||
1098 lHisSerGlnAsnIleAspPheGlyThrTyrAsnLeuThrIleLeuAsn 1114
      |||
3142 AAAGAC.....GCCGATTTCG 3158
      |||
      |||
      |||
1115 SerAspValIleLeuGlyGlyThrThrAlaIleAsnGlyGluIleAs 1131
      |||
3159 CCTGCATTAATCCGTCAAAGAACAGACCTTTCGACAACTCGGCAAG 3208
      |||
      |||
      |||
1131 pLeu.....LeuThrAsnAsnLeu..... 1137
      |||
3209 CAGAACCAAAAGACAGCGGCAAAAGAACAGCGCAACGCTTAGCGC 3258
      |||
      |||
      |||
1137 ..... 1137
      |||
3259 CTGATTGCGCGCGCGCGCGATGCCGTGCAAAAGACAGAAAGCTTCCGA 3308
      |||
      |||
      |||
1138 IlePheAlaAsnGlyThrSerThrTrpGlyAsnAsnThrSerLeuSerTh 1154
      |||
3309 ACCGCGCCGCGCAGCGCGCGGGAATGTCGCG...ATTATGACAGCGC 3355
      |||
      |||
      |||
1154 rThrLeuAsnValSerAsnGly...AsnValGlyGlnIleValIleAlaG 1170
      |||
3356 AGGAAAGAAAGAAAGGTCAGGCGGATTAAGACACCGCTTGGCGAAA 3405
      |||
      |||
      |||
1170 LuGlyAlaGlnValAsnAlaThrThrGlyThrThrIleLysIle 1186
      |||
3406 CAGCGCAGACGCAAAACCGCG...CCGCTACACCGCGCTCCCGCGC 3452
      |||
      |||
      |||
1187 GlnAspAsnAlaAsnAlaAsnPheSerGlyThrGlnThrTyrThrLeuI 1203
      |||
3453 CCGCGCGCGCGCGGATTTGCCGCACTGCAACCCCAACCGCAGCCCC 3502
      |||
      |||
      |||
1203 eGlnGlyGlyAlaArgPheAsnGlyThrLeu..... 1213
      |||
3503 AACCGCAGCGGACCTGATACCGCTTATTCGCAATAGCGTTTGAATGA 3552
      |||
      |||
      |||
1214 .....GlyAlaProAsn 1217
      |||
3553 TTTTCGCGACAGCTCACAGCGTTTTCGCGTACAGAGCAATTAAGCG 3602
      |||
      |||
      |||
1218 PheAspValThrGlyAsnAsnIlePhe...ValLysTyrIleLeu..... 1231
      |||
3603 GGTATTTCCGAGAGACCGCGCAAGCCGTTTGACACAGCGGATCCGG 3652
      |||
      |||
      |||
1232 .....IleArg 1234
      |||
3653 ACACCAAAACACTACCGTTTCGCAAGATTTTCGCGCTACCGCAACAAAC 3702
      |||
      |||
      |||
1234 sPAla.....AsnGlnAspTyr..... 1239
      |||
3703 GACCTGCGCAAAATCGTATGCAAGAAAACCTCGCAGCGCGCGCTCG 3752
      |||
      |||
      |||
1239 ..... 1239
      |||
3753 CATCTCTGTTTTCGCAACCGGACCGGAACACCTTCAGACGAGCATCG 3802
      |||
      |||
      |||
1240 .ValLeuThrArgThrAsnAspValLeuAsnValValThrThrAlaValG 1256
      |||
      |||
      |||
3803 GCAACTCGCAGCGCTTCGCCAGCGCGCG.....TTTTCGGCAATAC 3846
      |||
      |||
      |||
1256 LysAsnSerAla...IleAlaAsn..AlaProGlyValHisGlnAsnIleAl 1271
      |||
      |||
      |||
3847 GGCATCGACAGGTTCATACCTCGCATCAGCGCGCGGCTTTATAGAG 3896
      |||
      |||
      |||
1271 aIleCysLeuGluSerThrAspThrAlaAlaTyrAsnMetLeuAla 1288
      |||
      |||
      |||
3897 CG.....GCAGCTTTCAGACGCGCATCGAGGCAAAA 3928
      |||
      |||
      |||

```

```

1288 lalysaspSerSeraspValAlathrPheIleGlyAlaIleAlathrAsp 1304
3929 TCCGCCCGCGCTGCTGATTCAGGATTCAGGACGATCCGCCCGCT 3978
      ::::::::::::::::::::
1305 ThrGlyAlaIleAlaValAla...ThrValAsnLeuAsnAspThrGlnLys... 1319
      ::::::::::::::::::::
3979 TTGGCGGATTCGGCATCGACCGACGATCGCGGACGCGCTATTTC... 4026
      ::::::::::::::::::::
1320 ThrGlnAspLeuLeuGlyAsnArg...LeuGlyAlaLeuArgTyrLeuLeu 1335
      ::::::::::::::::::::
4027 .....
1335 rAsnSerGluThrAlaAspValGlyGlySerGluThrGlyAlaValSerS 1352
      ::::::::::::::::::::
4034 AAGCGGATTCAGGCTACGAAACGTC..... 4059
      ::::::::::::::::::::
1352 SerGlyAspGluAlaIleAspGlnValSerTyrGlyValTyrPalLysPro 1368
      ::::::::::::::::::::
4060 .....AATATGCCACGCC....GCCCTTGA...TTCAA 4088
      ::::::::::::::::::::
1369 PheTyrAsnIleAlaGluGlnAspLysGlyGlyLeuAlaGlyTyrLys 1385
      ::::::::::::::::::::
4089 CCGCTACCGCGCGGATTAAGCAGATTTATTCATCAACGCGCGCAAC 4138
      ::::::::::::::::::::
1385 sAlaLysThrAlaGlyValAlaValGlyLeuAspThrLeuAlaAsnSP 1402
      ::::::::::::::::::::
4139 ACATTCATCAGCGCTTATTTAGCTGTCTATACCATGCGCGCTTCG 4188
      ::::::::::::::::::::
1402 snLeuMetIleGlyAlaAlaIleGlyIleThrLysThrAsp..... 1415
      ::::::::::::::::::::
4189 GCGAAAGTCCGACACGCGTCAATACGCGCTATTTGGCTAGGATTTCCG 4238
      ::::::::::::::::::::
1416 .....IleLysHisGlnAspTyrLys 1422
      ::::::::::::::::::::
4239 CAAACCGCGAGTCGCGATGCGCTTAACGCGCAATCAAGCTTTCA 4288
      ::::::::::::::::::::
1422 sLys.....GlyAspLysThrAspIleLysGlyLeuS 1433
      ::::::::::::::::::::
4289 CGGTGCTCCCTCCACGCTGCGCGCGCAAGCGCGCACTGAGAGCGCAA 4338
      ::::::::::::::::::::
1433 erPheSerLeuTyrGlyAlaGlnGlnLeuValLysAsnPhenAlaGln 1449
      ::::::::::::::::::::
4339 CACAGCGCG 4347
      ::::::::::::::::::::
1450 GlySerAla 1452

```

seq_name: SwissProt_40:AMYH_YEAST

seq_documentation_block:

```

ID AMYH_YEAST STANDARD; PRT; 1367 AA.
AC P08640; P08068;
DI 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucoamylase SI/S2 precursor (EC 3.2.1.3) (Gluca 1,4-alpha-
DE glucosidase) (1,4-alpha-D-glucan glucosylhydrolase).
GN STA1 OR STA2 OR MAL5 OR YIR019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagsis K., Jones M.,
RA Louis E., Iye G., Moule S., Moule T., S. Hunt S., Jagsis K., Jones M.,
RA Rajandream M.A., Riles L., Rowley N., Skellon J., Pearson D.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RP [2]
SEQUENCE OF 1-242 AND 762-1331 FROM N.A.

```

```

RX MEDLINE=87194600; PubMed=3106330.
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STA1."
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Clavos M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STA2 and SGA genes
RT from Saccharomyces cerevisiae."
RL FEBS Lett. 239:179-184(1988).
CC -1 CAPALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1 SIMILARITY: TO S.POMBE SPEC215.13.
CC -1 SIMILARITY: SOME, TO S.POMBE SPEC285.13C.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL: Z38061; CAA86176.1; -
DR EMBL: M16164; AAA35014.1; -
DR EMBL: M16165; AAA35015.1; -
DR EMBL: X13857; CAA32069.1; -
DR PIR: B26877; B26877.
DR PIR: A26877; A26877.
DR SCD: S46478; S48478.
DR SGD: S0001458; MOC1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1367
FT DOMAIN 210 1367
FT CARBOHYD 817 817
FT CARBOHYD 874 874
SQ SEQUENCE 1367 AA; 136110 MW; 91C0062DBD61AA9D CRC64;

```

alignment_scores:

```

Quality: 232.50 Length: 1098
Ratio: 0.450 Gaps: 46
Percent Similarity: 47.086 Percent Identity: 20.036

```

alignment_block:

US-09-303-518d-649 x AMYH_YEAST

Align seg 1/1 to: AMYH_YEAST from: 1 to: 1367

```

1186 TCCCTTATTGACGAGAAAGCGATTTGATCTTACACGACATCA 1235
      ::::::::::::::::::::
25 AlaLeuValAlaProArgGlySerSerGluGlyThrSerGlyAsnSerIleVal 41
      ::::::::::::::::::::
1236 TCAAGCTGCTGAGGATTTATTTCCAGAGAGATTTACGCTCGCGCT 1285
      ::::::::::::::::::::
41 AsnGlyLysProAsnLeu.....AspPheAsnTyrPheIleMet 54
      ::::::::::::::::::::
1286 AAAATACGAAACTTGGCAAGCGCGCGCTGATATCAATCAAGACAGT 1335
      ::::::::::::::::::::
54 spGlnGlnAsnIleMetCln.....TyrThrLeuAspValThr 66
      ::::::::::::::::::::
1336 ACCGTACTTGGAAAGTAAAGCGCGTGGCAAGACCGCTGTCGCAAAAT 1385
      ::::::::::::::::::::
67 SerValSerTyrP.....ValGlnAspAsnThrTyrGlnIle 78
      ::::::::::::::::::::
1386 CCGCAAGCGACGCTGACGCTTCAAGCAAGGCGCAAGACGCTCGA 1435
      ::::::::::::::::::::

```

```
78 e.....ThrIleHisValIysGlyLys...GluAsn..... 87
1436 TCAGCGTGGCGGACGTAACGCTATTTGGATCAGCAGCGACGATATA 1485
88 ..... 111
1486 GGCATAAAACACAGCCTTAGTAATGCGCTTGTCAGCGGACGGGTAC 1535
91 LysTyrLeuTyrPseLeuLysIleIleGlyValThrGlyProLysGlyTh 1570
1536 GGTGCACTGTAATGCCGATATCACTTCAACC.....CCG 1570
107 rValGlnLeuTyrGlyTyrAsnGlu.AsnThrTyrLeuIleAspAsnPro 123
1571 ACNAACTGATTCGGCTTC.....GC 1593
124 ThrAspPheThrAlaThrPheGluValTyrAlaThrGlnAspValAsnSe 140
1594 GCGGACGCTTGGATTAAACGGGCATTCGCTTGTTCACCGCTATTCA 1643
140 rCysGlnValTyr..MetProAsnPheGlnIleGlnPheGlnTyrLeuG1 156
1644 AANTACCGATGAAGGCGCATGTTGTCAACCAATCAAGCAAGAAT 1693
156 nglySerAlaAla.....GlnTyrAlaSerS 165
1694 CCACCGTTACCATTCAGGCATTAAGATATTGTCTACAACCGCATTAAC 1743
165 erThrPglntPrGlyThrThrSerPheAspLeuSerThrGlyCysAsnAsn 181
1744 AACAGCTTGATAGCAAAAAAGAAATGGCTACACGGGTGGTGGGCA 1793
182 TyrAspAsnGlnGlyHisSerGlnThrAspPheProGlyPheTyrTrpAs 198
1794 GAAAGAT.....ACGACCA 1807
198 nIleAspCysAspAsnAsnCysGlyGlyThrLysSerSerThrThrThrs 215
1808 AACAGCAAGGGCGCTACACCTGTTTACAGCCCGCCGACAGAAAGCCGC 1857
215 erSerThrSerGlnSerSerThrThrThrSerThrSerThrSerLysSer 231
1858 ACCGTGCTGTTCCGCGGACAAATTTAAACGGCAACATCAACGCAAAC 1907
231 rThrThrThrSerThrSerThrSerThrThrThrThrSerThrThrs 248
1908 AACGCGCAACTGTTTTCAGCGGACACACCAACCCGACCGCTACAAATC 1957
248 erGlnSerSerThrSerSerSerThrThrAlaProAlaThrProThr... 263
1958 ATTAAAGGACGATTTGTCGCAAAAAGAGGCGCATTTCTCGCGGGAAATC 2007
264 .....ThrThrSerCysThrLysGluLys.....Proth 273
2008 GTGTGGGACAAACGACTGATCAACCGCACATTTAAAGCGGAAAACTTCA 2057
273 rProProThrThrThrSerCysThrLysGlu...LysPro..... 285
2058 AATTAAAGCGGACAGAGGGGTGTTCCCGCAATGTTGCCAAAGTGAAG 2107
285 ..... 285
2108 GCGATTGGCATTTGAGCAATCAGCCCAAGCATTTTGGTGTGCGACCG 2157
286 .....ThrProPro 288
2158 CATCAAGCGACACATCTGTACAGTTGCGACGCTGACGGGTGACAA 2207
289 HisHisAspThrThrProCysThrLysLysLysThrThrThrSerLysTh 305
2208 TTGTGTGCAAAAAACCATTCACGACAGATTAAGTGAATGCTTCATGACTA 2257
305 rCysThrLysLysThrThrThrProValProThrProSerSerSerThrTh 322

2258 AGACCGACATCAGCGGCATGTGATCTTCCGATCAGCGCTATTAAAT 2307
322 hrGlnSerSerSerAlaProValPro..... 330
2308 CTCACAGGCGTTGGCACATCAACGGGCATTTAGTGAATGGCGATAC 2357
330 ..... 330
2358 ACGTTATACAGTCAAGCCACAGCCACCAAAAGCGCAACCTTAGCCTCG 2407
331 .....ThrProSerSerSerThrThrGlnSer..... 339
2408 TGGGCAATGCCCAAGACATTTAATCAAGCCACATTTAAAGCCCAACACA 2457
340 .....SerSerAlaProValThrSerSerThr 348
2458 TCGGCTTGGCGCATGCTTCAATTATATCAAGCACACCGCCGACAA 2507
349 ThrGlnSerSerSerAla.....ProValProth 358
2508 CGGACGTCTGACGCTTTCGGCAACGCTAAGCAAGTAAGCCATTCCG 2557
358 rProSerSerSerThrThrGlnSerSerSerAlaProValThrSerSer 375
2558 CACTCAACGCGTAATGTCTCCCTAGCCGATTAAGGACGATTTCCATTGAA 2607
375 hrThrGlnSerSer..... 379
2608 AGACGCGCGTTTACCGGACAAATCAGCGCGGCAAGATTAAGGATTA 2657
380 SerAlaProValThrSerSerThrThrGlnSerSerSerAlaProValPr 396
2658 CTTAAAGACAGCAATGACGCTGCGCTCAGGACGACGATTAAGCAATT 2707
396 oThrProSerSerSerThrThrGlnSerSerSerSerAlaProValThrSers 413
2708 TAAACCTGACAGCGCACCATTAACATTCGCTTACCGCCACGAT 2757
413 erThrThrGlnSerSerSerSerAlaProValThrSerSerThrThrGlnSer 429
2758 GCGGACAGGCGGCGCAACCGGACGTGCGACAGATGCGCGCGCGC.... 2802
430 SerSerAlaProValThrSerSerThrThrGlnSerSerSerAlaProVa 446
2803 .CGTTCCGCGCGCTTCCGCGCGTCCCTAATTATCCGTACACGCGCACTT 2851
446 lThrSerSerThrThrGlnSerSerSerSerAlaProValThrProthSers 463
2852 CGGTAGAATCCGTTTCAACACGCTGACGGTAAACGGCAATTTGAACGCT 2901
463 erSerThrThrGlnSerSerSerSerAlaProVal.Thr..... 474
2902 CAGGCAACATTCGCTTATGTGAGACCTCTGGGCTACCGCAGCA 2951
475 .....SerSerThrThrGlnSerSerSerSerAlaProValProthr. 487
2952 ATTGAAGCTGGCGAAAGTTCCGAAGGACCTTACACCTTGCGGCTACACA 3001
488 .....ProSerSerSerThrThrGlnSerSerSer 497
3002 ATACCGGACAGCAACCTGCAAGCCTCGAACATTTGACGATGAGAGA 3051
498 AlaProValThrSer..... 502
3052 AAGAGCAACAAACCGCTGCGAAACCTTAATTTACACCTGCAAAACGA 3101
503 .SerThrThrGlnSerSerSerSerAlaProValProthProSerSerSer 519
3102 ACACG...TCGATGCCGGCGCTGCGGTACCACTCATTCGCAAAACG 3148
519 hrThrGlnSerSerSerAlaProAlaProthProSerSerSerThrThr 535
```

```

3149 GCGAGTCCGCGCTGATATCCGTCAGAAAGACAGAGCTTTCGACAA 3198
      |||:||||:||||:||||:||||:||||:||||:||||:||||:
536 GluSerSerSerAlaProValThrSerSerThr.....ThrGI 548
3199 CTCGGCAAGGAGAGAGCCAAAAAGAGCGGAAAAAGACAGCGCAAG 3248
      |||:||||:||||:||||:||||:||||:||||:||||:
548 UserSerSerAlaProValProThrProSerSerSerThrThrGluSers 565
3249 CATTGACGCGCTGATTCGGCGGCGGCGATGCGGTGCAAGAAACAGAAA 3298
      :|||:||||:||||:||||:||||:||||:||||:||||:
565 erSerThr.....ProValThrSerSerThrThrGluSerSer 577
3299 GCGTTCGCGCAACCGCGCCGAGCGAGCGGGAATAATGTCGCAATTATG 3348
      :|||:||||:||||:||||:||||:||||:||||:||||:
578 SerAlaPro.....ValProThrPro..... 580
3349 CAGCGGAGAGAGAGAAAAACGGGTGCGAGCGGATTAAGACACCGCCTT 3398
      :|||:||||:||||:||||:||||:||||:||||:||||:
581 .....ValProThrPro..... 584
3399 GCGCAAAACAGCGGAGAGCGGAAACCGCGCGCTACACCGCGCTCCGCC 3448
      :|||:||||:||||:||||:||||:||||:||||:||||:
585 .....SerSerSerThrThrGluSerSerSerAlaProValProThrPro 599
3449 GCGCGCGCGCGCGCGCGCGCGGATTTCCGCAACTGCAACCCGCAACGCGAG 3498
      :|||:||||:||||:||||:||||:||||:||||:||||:
600 SerSerSerThrThrGluSer.....SerSerAlaProAlaProThr 613
3499 CCCCACACCGAGCGCGACCTGATACGCGCTTATGCCAATAGCGGTTGAG 3548
      :|||:||||:||||:||||:||||:||||:||||:||||:
613 rProSerSerSerThrThrGluSerSerSerAlaPro.....V 626
3549 TGAATTTTCGCGCAGCGCTCAACAGCGTTTTCGCGGTACAGAGCAATTAG 3598
      :|||:||||:||||:||||:||||:||||:||||:||||:
626 alThrSerSerThrThrGluSerSerSerAlaProAlaProThr..... 640
3599 ACCGCGTATTTCGCGAAGACCGCGCAACCGCGTTTGACAGCAAGCGCATC 3648
      :|||:||||:||||:||||:||||:||||:||||:||||:
641 .....ProSerSerSerThrThrGluSerSerSerAlaProAlaPro 653
3649 CCGGAGACCA...AAGACTACCGTTGCGAAGATTTCCGCGCTACCGCGCA 3695
      :|||:||||:||||:||||:||||:||||:||||:||||:
653 lProThrProSerSerSerThrThrGluSerSerSerAlaProAlaPro 668
3696 ACAAAACCGCGCGCGCAATCGGATACAGAAAAACCTCGGACGCGCGC 3745
      :|||:||||:||||:||||:||||:||||:||||:||||:
669 .....ProThrProSerSerSerThrThrGluSerSerSerAlaProAla 683
3746 GCGTGGGATCCTGTTTCGACAAACGCGAGCAAAACACCTTCGAGCAG 3795
      :|||:||||:||||:||||:||||:||||:||||:||||:
684 ThrSerSerThrThrGluSerSerSerAlaProAlaThrSerSerThr 700
3796 GGCATCGGCAACTCGGACGCGCTTCCGACGCGCGCTTTCGGGCAATA 3845
      :|||:||||:||||:||||:||||:||||:||||:||||:
700 rGluSerSerSerAla...ProValProThr...ProSerSerSerThr 715
3846 CGGCATCG.....ACAGGTTCATATCGGCA 3871
      :|||:||||:||||:||||:||||:||||:||||:||||:
715 hrGluSerSerSerAlaProValProThrProSerSerSerThrThrGlu 731
3872 TCACCGCGCGGCGGCGGTTTACGACGCGGACCTTTCAGACGAGCATCGCA 3921
      :|||:||||:||||:||||:||||:||||:||||:||||:
732 SerSerSerAlaProValProThrProSerSerSerThrThrGluSer 747
3922 GCGCAAAATCGCGCGCGCTGTCGATTCAGGATTCAGGACAGCATACCG 3971
      :|||:||||:||||:||||:||||:||||:||||:||||:
748 .....SerSerAlaProValThrSerSerThrThrGluSerSerAla 762
3972 GCGCGGTTTGGCGGATTCGCGATTCAGAACGCGCATCGCGGCAACCGCT 4021
      :|||:||||:||||:||||:||||:||||:||||:||||:
762 lAlaProValProThrProSerSerSerThrThrGluSerSerAlaPro 778
4022 ATTTCGTCAAAAAGCGGATTACCGCTACGAAACGTCATATCGCACCC 4071

```

```

4118 .....ATTGATTCAAACCGCGGCAACACATTCCTCCATCCAGCCCTT... 4156
811 hrProSerSerSerSerAlaThrSerSerAlaProValProThrPro 827
4157 .....ATTGAGCGCTGTCATACCGATCCGCTTCG 4188
828 PheSerSerSerThrThrGluSerSerSerValProValProThrProSer 844
4189 GCGCAAACTCGCAACACGCGTCATACCGCGCTTATTCGTCAGAGATTTCG 4238
844 rSerThrThrGluSerSerSerSerAlaPro.....ValSers 856
4239 CAAACCCGCGAGTCGGAATGGGCGC.....TAAACCGCAAAATCAAG 4282
856 erSerThrThrGluSerSerValAlaProValProThrProSerSer 872
4283 .....GTTTCACGCTGCTCCCTCCAGCGTCCG 4309
873 SerAlaThrSerSerAlaProSerSerThrPro 884

seq_name: SwissProt_40:OMP_RICPR
seq_documentation_block:
ID OMP_RICPR STANDARD; PRT; 1643 AA.
AC 053020; G9ZCM0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (166 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas5) (rOmpB)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein omp); 32 kDa beta peptide].
CM OMPB OR SPAP OR SPA OR RP704.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=BREINL.
RX MEDLINE=91045972; PubMed=2122457;
RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
RT "Characterization of the gene encoding the protective paracrystalline-
RT surface-layer protein of Rickettsia prowazekii: presence of a
RT truncated identical homolog in Rickettsia typhi."
RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BREINL.
RA Moron C.G., Yu X.J., Walker D.H.;
RT "Sequence analysis of ompB of Rickettsia prowazekii."
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Scheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).
RN [4]
RP PARTIAL SEQUENCE.
RC STRAIN=BREINL.
RX MEDLINE=92114896; PubMed=1370573;

```


1500 CTTTACGAAATCGGCTTGTGTCAGCGGACGGGTACGGTG...CACTGA 1546
1534 aleuGlnHisIleThrLeuAlaAsnAspAlaSerLysIleLeuAlaLeu 551
1547 ATGCCGATATGATGTTCAACCCCGAC.....AACTCTATTTCGGC 1587
551 sp61yAlaAsnIleIleGlyAlaAsnValIleGlyAlaIleHisPheGln 567
1588 TTTCCGGGGGACGTTTGGATTTAACGGGCACTTCGCTTCGTCACGG 1637
568 AlaAsnGlyGlyThrIle.....Lys 574
1638 TATTCAAAATAACGATGAAGGGGAGATGATGTCACCAACATCAAGACA 1687
574 sleuthrAsnThrGlnAsnAsnIleVal.....AsnPheAspL 588
1688 AAGAAATCACCGCTTACCTTACAGGCAATAAAGATATTGCTACACCGCG 1737
588 euAspIleThrAspLysThrGlyValValAlaAspAlaSerLeuThr 604
1738 AATAACAACAGCTTGATAGCAAAAAGAAATGGCTACACCGCTTGCTT 1787
605 AsnAsnGlnThrLeu.....ThrIleAsnGlySerIle 615
1788 TGGCGAG...AAAGATACGACAAACGAGCGGCGCTCAACCTTGTT 1834
615 eGlyThrValAlaAlaAsnThrLysThrLeuAlaGlnLeuAsnIle... 630
1835 ACCGACCCCGGAGAACCGGACCGCTGCTTCCGCGGACAAAT 1884
631GlySerSerLysThrIleLeuAsnAlaGlyAspValAla 643
1885 TTAAACGGCAACATCAACGCAAAACGCAACCTGTTTTCAGCGGACG 1934
644 IleAsnIleLeuValIleGlnAsnAsnGlySerVal..... 655
1935 ACCAACCGCGACGCTACATGATTTAAACGACCTTGTGTCGCAAAA 1983
656GlnLeuAsnHis...AsnThrLysLeuIleThrLysT 667
1984GAGGCAATTCCTCGCGGGAATGCTGTGGACACGACTGAT 2028
667 hrIleAsnAlaAlaAsnGlnGlyIleIleValAlaAlaAspProLeu 683
2029 AACCGCACATTT.....AACGCGAAAA 2051
684 AsnThrAsnThrThrLeuAlaAspGlyThrAsnLeuGlySerAlaGluAs 700
2052 C.....TTCCAAATTAAGCGCGACGCGGCGGTGTT 2083
700 nProLeuSerThrIleHisPheAlaThrLysAlaAlaAsnAlaAspSerI 717
2084 CCGCAATGTTCGCAAGGAAAGCGGATTTGGCAATTCAGCAATCAGCC 2133
717 leuAsnValIleLys..... 722
2134 CAAGCAATTTTGTGTGTCGACCGCATCAAGCCACAAATCTGTACAG 2183
722 722
2184 TTCGACATCGACGCGGTGACAAATGTGTGCAAAAAACATTCACGACG 2233
723GlyValAsnLeuIleValAlaAsnAsnIleThrThrAsnA 735
2234 ATAAAGTATGCTTCATTTG.....ACTAGACCGGACATC 2268
735 spAlaAsnValIleGlySerLeuHisPheArgSerGlyThrSerIleVal 751
2269 AGCGGC.....AATGTCGATCT 2285
752 SerGlyThrValIleGlyGlnGlnGlyHisLysLeuAsnAsnLeuIleLe 768

2286 TGCCGATCAGCGCTCATTTAAATCTCACAGGCGTTCACACTACAGCGCA 2335
768 uAspAsnGlnThrThrValLysPheLeuGlyAspThrThrPheAsnIleG 785
2336 ATCTTAGTGCAAATGCG.....GATACACGCTTATACA 2367
785 LyThrLysIleGlnGlyLysSerIleLeuGlnIleSerAsnAsnIleThr 801
2368 GTCACGCAAC.....AACGCCACCAAAACCGCAACTTACGCTGCGG 2411
802 ThrAspHisValIleGlySerAlaAspAsnThrGlyThrLeuGlnPheValAs 818
2412 CAATGCCCAACCAACATTTATCAAGCCACATTAACGCGAAC..... 2454
818 nThrAspProIleThr.....ValThrLeuAsnLysGlnGlyAlaI 832
2455ACATGCGCTTCGGCAATGCTTCA 2478
832 yrpPheGlyValLeuLysGlnValIleIleSerGlyProGlyAsnIleVal 848
2479 TTTAAT.....CTAAGCA 2492
849 PheAsnGlnIleGlyAsnValIleValHisGlyIleAlaAlaAsnSe 865
2493 CCACGCCGTACAAACCGGACGCTGACGCTTTC..... 2526
865 rIleSerPheGlnAsnAlaSerLeuGlyThrSerLeuPheLeuProSerG 882
2527GGCAACGCTAAGCAACGTAAGCCATTCGCGCA 2559
882 LyThrProLeuAspValLeuThrIleLysSerThrValIleLysAsnGlyThr 898
2560 CTCACGCGTATGTCCTCCATAGCGGATTAAGCGATTTTCATTTGAAG 2609
899 ValAsp...AsnPheAsnAlaProIleValValValIleSerGlyIleAspSe 914
2610 CACCGCGTTTACCGGACAAATCAAGCGCGGCAAGATACGCAATACAT 2659
914 MetIleAsnAsnGlyGlnIleIleGlyAspLysLysAsnIleIleAlaI 931
2660 TA.....AAGCACGCGAATGAGCGCTGCGCGGACGCAATTA 2700
931 euSerLeuGlySerAspAsnSerIleThrValAlaAlaAsnThrLeuIle 947
2701 GGCAATTTAACTT.....GACACGCGACCATTAACATTCATTC 2741
948 SerGlyIleArgThrThrLysAsnAsnGlnGlyThrValThrLeuSerG 964
2742 CGCGTATCGCGACGAT.....GGCGACAGGCGGCAACCG 2776
964 yGlyMetProAsnAsnProGlyThrIleTyrGlyLeuGlyLeuGlnG 981
2777 GCAGTCCGACAGATCGCGCGCGCGCTTCGCGCGCTTCGCGCGCTTCC 2826
981 LysPro.....Lys 984
2827 CTATATTCGTTTACCGCGCACTTCGCTAGATCCCGT..... 2865
985 LeuLysGlnValThrPheThrThrAspLysAsnAsnLeuGlySerIleI 1001
2866 TTTCAACGCGGTGACGTTAAC..... 2886
1001 eAlaAsnAsnValThrIleAsnAspTyrValThrLeuThrGlyGly 1018
2887GGCAATTTGAACGCT 2901
1018 leaIleGlyThrAspPheAspAlaLysIleThrLeuGlnGlySerValAsnGly 1034
2902 CAGGCAACATTCGCTTATGTCGGAACCTTCGCGTACGCGACGACGA 2951
1035 AsnAlaAsnValArgPheValAspSerThrPheSerAspProArgSer 1051
2952 ATTGAAGCTGGCGGAAGTTCCGAGGCACTTTCACCC.....T 2989

```

1051 ttleValaIaIaThrGlnAlaAsnLysGlyThrValThrTyrLeuGlyAsnA 1068
      ::::::::::::::::::::|||
2990 TGGCGGTCAACAATACCGC.....AACGAACCTCGCAAGCCTGACAA 3033
      |||::|
1068 IaleuValSerAsnIleGlySerLeuAspThrProValAlaSerValArg 1084
      |||::|
3034 TTGACGGTGTGGAAGAAAGCAACAACACGCTGTCCGAAACCTTAA 3083
      |||::|
1085 PheThr.....GlyAsnAspSer.....GlyAlaGlyLeuG 1095
      |||::|
3084 TTTCACCTTCGCAAAACGAACACGTGATCCGCGCGCGGTACAC 3133
      ::::::::::::::::::::|||
1095 nGlyAsnIleTyrSerGlnAsnIleAspHeGlyThrTyrAsnLeuThr 1112
      ::::::::::::::::::::|||
3134 TCATCCGCAAAAGACGGCGAGTTCCGCTCATATCCGCTCAAGACAA 3183
      ::::::::::::::::::::|||
1112 IeLeuAsnSer.....AsnValIle..... 1118
      ::::::::::::::::::::|||
3184 GACCTTTCGCAAACTCGGCAAGCGAAGCCAAAACAGCGGAA 3233
      |||::|
1119 .....LeuGlyGlyThrThrAlaIleAsnGlyGlu 1130
      ::::::::::::::::::::|||
3234 AGACACGCGCAACCTTGACCGCGTGTGCGCGCGCGGATCCG 3283
      ::::::::::::::::::::|||
1130 eAspLeuLeuThrAsn...AsnLeuIlePheAlaAsnGlyThrSerThr 1146
      ::::::::::::::::::::|||
3284 TCGAAAGACAGAAAGCGTTCGCAACCGCGCGCGAGCGCGGAA 3333
      ::::::::::::::::::::|||
1146 rpgIyAspAsnThrSerIleSerThrThrLeuAsnValSerSergly 1161
      ::::::::::::::::::::|||
3334 AATGTCGCG...ATTATGACGCGGAGAGAAAGAAAAACGGGTGACGC 3380
      |||::|
1162 AsnIleGlyGlnValIleAlaGlnAspAlaGlnValAsnAlaThrThr 1178
      ::::::::::::::::::::|||
3381 GGATTAAGACACCGCGCTTGCGAAACAGCCGAAACCGGAAACCCG...C 3427
      ::::::::::::::::::::|||
1178 rThrGlyThrThrThrIleIleIleIleIleIleIleIleIleIleIle 1195
      ::::::::::::::::::::|||
3428 CGGCTACACCGCGCTTCGCGCGCGCGCGCGCGCGGATTCGCG 3477
      |||::|
1195 eArgIleThrGlnAlaThrThrLeuIleGlnGlyAlaArgPheAsnGly 1211
      ::::::::::::::::::::|||
3478 CAACGTGCACACCCCAACCGCAACCGCAACCGGAGCGACCTGATCAGCG 3527
      |||
1212 ThrLeu..... 1213
      ::::::::::::::::::::|||
3528 TTATGCCAATAGCGGTTTGAAGTAATTTCCGCGACGCTACACGCTT 3577
      |||::|
1214 .....GlyAlaProAsnPheAlaValThrGlySerAsnIleP 1226
      ::::::::::::::::::::|||
3578 TCGCGGTACAGACGATTTAGACCGCGTATTTGCCGAAGCCGCGCAAC 3627
      |||::|
1226 he...ValIleTyrIleLeu..... 1231
      ::::::::::::::::::::|||
3628 GCCGTTTGGACAAGCGCATCCGGGACACCAACACTACCGTTGCGAGA 3677
      |||::|
1232 .....IleArgAspSer.....AsnGlnAs 1238
      ::::::::::::::::::::|||
3678 TTTCGCGCGCTACCGCCCAACAACCGACCTGCGCAATCGTATGAGA 3727
      |||
1238 pTyr..... 1239
      ::::::::::::::::::::|||
3728 AAAACCTCGGCAAGCGCGCGCGCTGCGCATCTTTTCGACACACGCGAC 3777
      |||::|
1240 .....ValLeuThrArgThrAsnAspVal 1247
      ::::::::::::::::::::|||
3778 GAAACACCTTCGACAGCGCATCGCAACTCGGACAGCGCTTCCGACAG 3827
      |||::|
1248 LeuAsnValValThrThrAlaValGlyAsnSerAla...IleAlaAsnA 1263
      ::::::::::::::::::::|||
3828 CGCGCTTTTCGGCATACGCGCATG...ACAGGTTTCATCATCGGCA 3871
      |||::|

```

```

1263 IapTrGlyValSerGlnAsnIleSerArgCysLeuGlnIleThrAsnThr 1279
      ::::::::::::::::::::|||
3872 TCAGCGCGCGCGCGGTTTACGACG.....CGACG 3903
      |||::|
1280 AlaAlaTyrAsnAsnMetLeuAlaIleLysAspProSerAspAlaThr 1296
      |||::|
3904 CTTTACAGCGCATCGAGCAAAATCCGCGCGCGCTGCTCATACGG 3953
      |||::|
1296 rHeValIleGlyAlaIleAlaThrAspThrSerAlaAla.....ValThrP 1311
      ::::::::::::::::::::|||
3954 CATTCAGCGCATACCGCGCGCGGTTTGGCGGATTCGACGACCG 4003
      ::::::::::::::::::::|||
1311 hrValAsnLeuAsnAspThrGlnIleIleIleIleIleIleIleIleIle 1327
      ::::::::::::::::::::|||
4004 ACATCGCGCAACCGCGCTATTCGTCACAAAGCGGATTCAGCTACGA 4053
      ::::::::::::::::::::|||
1328 ..LeuGlyThrLeuArgTyrIleSerAsnAla.....Glu 1338
      ::::::::::::::::::::|||
4054 AACGTCAATATCGCACACCCCGCTTGCAATCAACCGCTACCGCGGG 4103
      ::::::::::::::::::::|||
1339 ThrSerAspValAla.....GlySerAlaThrGlyAlaValSerSerg 1353
      ::::::::::::::::::::|||
4104 CATTAAGCGCATATTCATTCACAAACCGCGCAACATTCATCCATCAGC 4153
      ::::::::::::::::::::|||
1353 yAspGlnAlaGlnValSerTyr.....GlyValThrAlaLysP 1366
      ::::::::::::::::::::|||
4154 CTATTTTGAACCTGTCCTATACGATGCGCGCTTCGCGC..... 4191
      |||::|
1366 roHeTyrAsnIleAlaGlnIleAspLysGlyGlyIleAlaGlyThr 1382
      ::::::::::::::::::::|||
4192 AAAGTCCGCAACCGC..... 4206
      ::::::::::::::::::::|||
1383 LysAlaLysThrThrGlyAlaValAlaGlyLeuAspThrLeuAlaSerAs 1399
      ::::::::::::::::::::|||
4207 .....GTCAATACCGCGCTATTCG 4225
      ::::::::::::::::::::|||
1399 pAsnLeuMetIleGlyAlaAlaIleGlyIleThrLysThrAspIleLysH 1416
      ::::::::::::::::::::|||
4226 CTCGAGATTTGCGAAACCCGCAATGCGCGCTAAACGCGCA 4275
      |||::|
1416 IsgInAspTyrLys.....GlyAspLysAsp 1426
      ::::::::::::::::::::|||
4276 ATCAAGGTTTACGCTGCTCCACGATGCGCGCGCGCAAGCGCGCA 4325
      |||::|
1427 IleAsnGlyLeuSerPheSerIleTyrGlySerGlnGlnLeuValLysAs 1443
      ::::::::::::::::::::|||
4326 ACTGAAAGCGCAACACAGCGCG 4347
      ::::::::::::::::::::|||
1443 nPhePheAlaGlnGlyAsnAla 1450
      ::::::::::::::::::::|||

```

seq_name: SwissProt_40:HIVA_PROMI

seq_documentation_block:

| ID | HLVA_PROMI | STANDARD: | PRT: | 1577 AA. |
|-----------------|---|-----------|------|----------|
| AC | P16466; | | | |
| DT | 01-AUG-1990 (Rel. 15, Created) | | | |
| DT | 01-AUG-1990 (Rel. 15, Last sequence update) | | | |
| DT | 01-NOV-1990 (Rel. 16, Last annotation update) | | | |
| DE | Hemolysin precursor. | | | |
| GN | HPWA. | | | |
| OS | Proteus mirabilis. | | | |
| OC | Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; | | | |
| CC | Proteus. | | | |
| NCBI_TaxID=584; | | | | |
| OK | [1] | | | |
| RN | SEQUENCE FROM N.A. AND SEQUENCE OF 30-43. | | | |
| RP | STRAIN-ISOLATE 477-12; | | | |
| RC | MEDLINE=90170827; Pubmed=2407716; | | | |
| RX | Upjohn T.S., Welch R.A.; | | | |
| RT | "Nucleotide sequencing of the Proteus mirabilis calcium-independent hemolysin genes (hpmA and hpmB) reveals sequence similarity with the Serratia marcescens hemolysin genes (shlA and shlB)."; | | | |
| RT | Serratia marcescens hemolysin genes (shlA and shlB)."; | | | |
| J. Bacteriol. | 172:1206-1216(1990). | | | |


```

427 AsnLysAspIleHisIleasnGlyLeuValGlnLysGluSerArgSerG 443
      ||| ||| ||| ||| |||
428 AATGGGAATTAAGTCTTTAAGCAGCATTAATGACAGCAAGAAATCA 1027
      ||||| ||||| ||||| |||||
443 uasnGly.....AsnLysArgAsnHisThrSerArgLeu. 454
      ||||| ||||| ||||| |||||
1028 ATGCCAATCATGAACAATTTCTGCTTAATAGATTAAACAGCAAGC 1077
      ||| ||||| ||||| |||||
455 .....GluSerGlySerTrpSerAsnSerHisGlnThrGluThr 467
      ||| ||||| ||||| |||||
1078 GTTCAATGTTTAATGTTCTTAA..... 1101
      ||||| ||||| ||||| |||||
468 LeuLysAlaSerGlnLeuThrAlaGlyLysAspLeuGlyLeuAspAlaG 484
      ||||| ||||| ||||| |||||
1102 ....TCCGACAGCAGCAAGAACCTGTTTATCATGCT..... 1134
      ||| ||||| ||||| |||||
484 nGlySerIleThrAlaGlnGlyAlaLysLeuHisAlaAsnGlnVal 501
      ||||| ||||| ||||| |||||
1134 ..... 1134
501 euValAsnAlaLysAspAsnIleAsnLeuAsnValGlnLysThrAsnAsn 517
      ||||| ||||| ||||| |||||
1135 .....GCAGTGG 1142
518 AspLysThrValThrAspAsnHisValMetTrpGlyLysIleGlyLys 534
      ||||| ||||| ||||| |||||
1143 TGTCAACAGTTATCGACCCAGCATTAATGAGAAATATTTCTCTTA 1192
      ||||| ||||| ||||| |||||
534 yGlnAsnLysAsnAsn.....AsnAsnGlnGlnGlnValSerHisA 548
      ||||| ||||| ||||| |||||
1193 TTGACGAA.....GGAAGAGCGAATGATTAATCCAGCAACATCAT 1236
      ||||| ||||| ||||| |||||
548 lathGlnLeuThrAlaAspGlyGlnLeuLeuAlaAlaAspAsnAsn 564
      ||||| ||||| ||||| |||||
1237 CAAGGTGTGAGAGATTATTTCCAGAGATTTTACGGTCTCCCTGA 1286
      ||||| ||||| ||||| |||||
565 ValAsnIleThrGlySerGlnValLysGlyAsn..... 575
      ||||| ||||| ||||| |||||
1287 AATTAACGAATTTGGCAAGCGCGCGCTCATATACGTAGACAGTA 1336
      ||||| ||||| ||||| |||||
576 .....GlnGlyAlaPheValLysThrThrGlnGlyAsp 587
      ||||| ||||| ||||| |||||
1337 CCGTTACTTGAAGAATAAGCGCGTGGCAAGACCGCTGTCCAAATC 1386
      ||||| ||||| ||||| |||||
587 alVal.....IleAspAsnAlaLeuSerGlnThrIleSerLysIle 600
      ||||| ||||| ||||| |||||
1387 GGCAAA.....GGCAGCTGCAC.....GTTCAAGCCAAAGGGGAAA 1424
      ||||| ||||| ||||| |||||
601 AspIleuArgThrGlyThrAlaPheAsnIleThrLysSerSerHisLysAs 617
      ||||| ||||| ||||| |||||
1425 CCAAGGCTCGATCAGCGTGGCGACGATTCATTTGATCCAGCAG 1474
      ||||| ||||| ||||| |||||
617 nGlnThrAsnLysGlnThrSerThrLysSerGlnLeuIleSerAsp.... 632
      ||||| ||||| ||||| |||||
1475 CAGACGATAAGCGCAAAAAACAGCTTTAGTGAAATCGCTTGTCAGC 1524
      ||||| ||||| ||||| |||||
633 .....AlaGlnLeuThrValValSer 639
      ||||| ||||| ||||| |||||
1525 GGCAGG.....GTACGCTGCAACTGAATGCCGATATCA 1559
      ||||| ||||| ||||| |||||
640 GlyAsnAspValAsnValIleGlySerLeuLysSerAla..... 653
      ||||| ||||| ||||| |||||
1560 GTTCAACCCCGCAAACTCTATTTCGGCTTCGGCGGACGATTGGATT 1609
      ||||| ||||| ||||| |||||
654 .....AspLysLeuGlyIleHisSerLeuGly.....AspI 664
      ||||| ||||| ||||| |||||
1610 TAAACGCGATTCGCTT...TGTTCACACCGATTCAAATACCGATGAA 1656
      ||||| ||||| ||||| |||||
664 leAsnValLysSerAlaGlnGlnValThrLysIleAspAspGlnLysThr 680
      ||||| ||||| ||||| |||||
1657 GGGGCGATGATTTGTCAACCAATCATCAAGCAAAAGATCCACC..... 1698
      ||||| ||||| ||||| |||||

```

```

681 SerLeuAlaIleThrGlyHisAlaLysGlnValGlnAspLysGlnTyrSe 697
      ||||| ||||| ||||| |||||
1699 .....GTTACCATTAAGCGCAATTAAGATTTGCTGACA.... 1731
      ||||| ||||| ||||| |||||
697 rAlaGlyPheHisIleThrHisThrThrAsnLysAsnHisThrGlnGut 714
      ||||| ||||| ||||| |||||
1732 .....ACCGCAATTAACAGCTTGATAGC 1758
      ||||| ||||| ||||| |||||
714 hGlnGlnAlaAsnSerThrIleSerGlyAlaAsnValAspLeuGlnAla 730
      ||||| ||||| ||||| |||||
1759 AAAAAAGAAATTTCCCTAACCGCTTGCTTGGCGGAAAAAGATACGACCA 1808
      ||||| ||||| ||||| |||||
731 AsnLysAspValThrPheAlaGly.....SerAspLeuLysThrThrAl 745
      ||||| ||||| ||||| |||||
1809 AACCAAGCGGCGGCTC.....AACCTGTTTAC..... 1836
      ||||| ||||| ||||| |||||
745 aGlyAsnAlaSerIleThrGlyAspAsnValAlaPheValSerThrGln 762
      ||||| ||||| ||||| |||||
1837 ..CAGCCCGCGCGAAGACCGCACCTGCTTCCGCGGAAACCAAT 1884
      ||||| ||||| ||||| |||||
762 snLysLysGlnThrAspAsnThrAspThrIleSerGlyGlyPheSer 778
      ||||| ||||| ||||| |||||
1885 TTAAACGCAACATCAGCAACAAACGCAACTGTTTTCAGCGGCGAG 1934
      ||||| ||||| ||||| |||||
779 TyrThrGlyGlyValAspLysValGlySerLysAlaAspPheGln-TyrA 795
      ||||| ||||| ||||| |||||
1935 ACCAAGCAGCGCAGCTCAATCATTAATTAACGACCATTTGCTCCCAAAAG 1984
      ||||| ||||| ||||| |||||
795 sPlyGlnHisThrGlnThrGlnVal.....ThrLysAsnArgLysSer 809
      ||||| ||||| ||||| |||||
1985 AGGCGATTCCTCGCGGGAATTCGTGGGACAGCATGATCAACCGC 2034
      ||||| ||||| ||||| |||||
810 GlnThrGlnValAlaGlyAspLeuThrIleThrAlaAsnLysAspLeu 826
      ||||| ||||| ||||| |||||
2035 ACATTTAAAGCGGAAACTTCAATTAAGCGGACAGCGGCTTTC 2084
      ||||| ||||| ||||| |||||
826 uHis..GlnGlyAlaSerHisHisValGlnGlyArgTyrGlnGluSerG 842
      ||||| ||||| ||||| |||||
2085 CCGCAATGTTGCCAAGTGAAGGCGATTTGGCATTTGAGCAATCGCC 2134
      ||||| ||||| ||||| |||||
842 yGlnAsnIleGln.....HisLeu..... 848
      ||||| ||||| ||||| |||||
2135 AAGCACTTTTGGTGTGCGACCGCATCAAGCCAGACATCTGTACAGCT 2184
      ||||| ||||| ||||| |||||
849 .....AlaValAsnAspSerGlnThrSerLysThrAsp 859
      ||||| ||||| ||||| |||||
2185 TCG.....GACTGCGAGCGGTCT 2201
      ||||| ||||| ||||| |||||
860 SerLeuAsnValGlyLysAspValGlyValAsnLeuAspTyrSerGlyVa 876
      ||||| ||||| ||||| |||||
2202 GACAAATTTGTGGAAGAAACCATTAACGACGATAAGTGAATTCGTCAT 2251
      ||||| ||||| ||||| |||||
876 lThrLysProValLysLysAlaIleGlnAspLeu.....ValAsnT 890
      ||||| ||||| ||||| |||||
2252 TGACTAAG.....ACGACATCAAGCGGCAATGTGATCTTGCC 2289
      ||||| ||||| ||||| |||||
890 hThrLysProGlyAsnAsnThrAspLeuThrLysLysValThrAlaArg 906
      ||||| ||||| ||||| |||||
2290 GATCAC...GCTCATTTA...AATTCACAGGCTTGCCAGACATCAAC.. 2331
      ||||| ||||| ||||| |||||
907 AspAlaIleAlaAsnLeuAlaAsnLeuSerAsnLeuGlnThrProAsnVa 923
      ||||| ||||| ||||| |||||
2332 .....GGCAATCTTACTGCAATAGCCG 2353
      ||||| ||||| ||||| |||||
923 lGlyValGlnValGlyLysGlyGlySerGlnGlnSerGlnThrA 940
      ||||| ||||| ||||| |||||
2354 ATACAGCT...TATACAGTACGCCAAGCGC..... 2382
      ||||| ||||| ||||| |||||
940 sPserGlnAlaValSerThrSerIleAsnAlaGlyLysIleAspIleAsp 956
      ||||| ||||| ||||| |||||
2383 .....ACCCAAA 2390
      ||||| ||||| ||||| |||||
957 SerAsnAsnLysLeuHisAspGlnGlyThrHisTyrGlnSerThrGlnG 973
      ||||| ||||| ||||| |||||

```

```

2391 CGGCAACACCTTACCCCTGCGGCATCC..... 2418
973 ucll.....lleSerleuThrAlaSerThrHisThrSerGluAlaThrLeuA 989
2419 .....CAAGCAACATTTATATCAAGGCACATTTAAAGGCACACATCG 2460
989 sPLyHisGlnThrThrPheHisGluThrLysGlyGlyGlnIle60ly 1005
2461 GCTTGGGCAATGCTCTCATTTAATCTTAAGCAGCAGCCGACACAAAGG 2510
1006 ValSerThrLysThrGlySerAspIleThrValAlaIleLysGlyGlnI 1022
2511 CAGTCTGACGGCTTCCGGC.....AACGTAAGGCAAGGTAGCCATT 2554
1022 YglThrThrAspAsnAlaLeuMetGluThrLysAlaIleLysSerGln 1039
2555 CCGGACATCAACGGTAATGTCTCCCTACCGCATAGGCAAGTATTCATT 2604
1039 heThrSerAsnGlyAspIleSerThrLeuAsnValGlyIleAsnAlaHis 1055
2605 GAAGCAGCGGCTTACCGGCAA.....ATCAGCGCGG 2639
1056 GluGlyIleGlnPheAspAlaGlnLysGlyLysThrValIleAsnIleG 1072
2640 CAAGGATACGGCATTCACCTTAAAGAAGCAGTAAGCAGCTCCGCTCAG 2689
1072 YglAspLeuThrLeu..... 1077
2690 GCACGGAATTAGCATTAAACCTTGACACAGCCACCATTCACCTCAAT 2739
1077 ..... 1077
2740 TCGGCTATCCGACGAGATCGGCGAGGGCGCAACGGCAGTGGCAGCA 2789
1078 .....AlaGlnAlaThrAs 1082
2790 TCGCGCGCGCCGCGTTCGCGCGCTTCGCGCTCCCTATTATCCGTTA 2839
1082 P..... 1082
2840 CACGCCACACTTCGGTAGAATCCGTTTCAACACAGCTGACGTAAAGCG 2889
1083 .....ThrlSerGlnSerGlnSerAsn.....ValAsnIly 1093
2890 AAATGACACGGTCAGAGGAACATTCGCTTATGTGCGAATCTTCGCGCTA 2939
1094 SerAlaAsn..... 1096
2940 CCGCAGCAGCAAAATTTGAAGCTGGCGGAAAGTTCGAGAGCACTTACACT 2989
1097 .....LeuLysValGlyThrThrProGlnSerLysAspTyrG 1109
2990 TGGGGGTCAACATATCCGCGACAGAACCTCGAACCAATTATACG 3039
1109 LysGlyGlyPheAsnAlaGlyThrThrHisSerLysGlnGlnThrThr 1125
3040 .....GAGGAGGAAGGAAAGACAAACCGCTGCGCAAA 3077
1126 AlaLysValGlyThrIleThrGlySerGlnGlyIle60IleuAsnIleG 1142
3078 CCTTAATTTACCTCGCAAAAGCAGACAGCTGATGCCGCGCGTGCCTT 3127
1142 YHisAsnLeuThrThrLengGlnGlyThrHisLeuSerSer..... 1154
3128 ACCAAGCTCATCCGCAAAAGCGGGAATTCGCGCTGCATATATCCGCTC 3174
1155 .....GluGlnAspIleAlaLeuAsnAlaThrAsnLysValAsp 1167
3175 .....AAGAAGCAACGCTTTCCGA 3194
1168 LeuGlnSerAlaSerSerGlnHisThrGlnLysGlyAsnAsnLeuSerG 1184

```

3195 CAACACTCGGCAAGCGCAAGAACCCAAAAACAGCGCGGAAAAAGCAACAGCGCG 3244
 1184 yelValglInalagIlyrPheglYlyslMetThr.....AspAspAla 1199
 3245 AAAGCGTTGACGCGCTGATGTGGCGGGCGCGGATGCCGTGCAAAAGACA 3294
 1199 erSerValAsnGlyLeu...GlySerIAlaGlnPheAlaIleGlyLysGln 1214
 3295 GAAGCGTTGCCGAACCGCGCGCGGACGAGCG... 3327
 1215 AspGlu...LysSerValSerArgGlnGlyGlyThrIleAsnAsnSerGln 1220
 3328GGGAAATATCGGCATTATGACAGCGCGGAGAG 3361
 1220 yAsnLeuThrIleAsnGlnAsnSerValHisLysGlnGlyAlaGlnValAla 1247
 3362 AGAAAAACGGCGTCAGCGG... 3381
 1247 snSerLysAspThrGlnLeuThrSerGlnSerGlnAspIleGlnIleThr 1263
 3381 3381
 1264 SerAlaGlnSerThrAspTyrLysAsnAsnTyrGlyThrAspIleGlyPhe 1280
 3382GATAAAGACACGCGCTTTGGCGAAMACGCGCGAAGCGAAACCGCGG 3427
 1280 eaSnGlyLysLysThrAsnAsnThrProLysGlnValThrGlnGlnLysPhe 1297
 3428 CGGCTGACACCGCGCTTCCCGCGCGCGCGCGCGCGCGCGGATTTGGCG 3477
 1297 roIaIaThrSerIleHisAsnIleGlyGlyLysLeu.....IleVal 1310
 3478 CAACCTCAACCCCAACCGCGACGCCCAACCGACGCGAGCGACTGATCAGCG 3527
 1311 AsnValGlnAspGlnGlnIlePheThrSerHisGln..... 1321
 3528 TTATGCAATAGCGGGTTAGTGAATTTTCCCGACGCTCAACAGCGTTT 3577
 1322AsnAlaThrIleGlnGlyThrGlyThrIleAsnSer.... 1334
 3578 TCGCGGTACAGGACGATTTAGACCGCGATTTTGGCGAAGCGCGCGCAAC 3627
 1335AsnLysAspLeuThrIleSerGlyAlaAsnValThrAlaAsp 1348
 3628 GCCGTTTGACACAGCGGCGATCCGGACACCAACACTACGCTGCGCAAGA 3677
 1349 SerVal...ThrGlyAsnValGlyLysSerLeuHisIleAlaSerGln.. 1363
 3678 TTTCGCGCCTACGCGCAACAACCGACTGCGGCAATTCGGTATGCAAGA 3727
 1364LysGlnSerAsp...ArgHisVal..... 1370
 3728 AAACCTCGCACCGCGCGCGCTGCGCATCTGTTTGGCAACCGGACG 3777
 1371ThrValGlyValAsnValGlyTyrAsnHisThr 1381
 3778 GAAACACCTTGCAGACGGCGATCGGGAACTCGGACAGCGCTGCCACGG 3827
 1382 AsnAspProLysSerSerGlnValAsnLysThrAlaLys...AlaGlyGln 1397
 3828 CGCGCTTTTGGGCAATAGCGGATCGACAGTTCTACATGGATGACGCG 3877
 1397 ySerLeuLeuGlnGlyThrIleLysAspThrIleAspSerGlyIleLys 1414
 3878 CGGACGCGGGTTTTAGACG.....GGACGCTTTTACAGCGCATC 3918
 1414 erSerThrAspAlaIleSerAspLysTyrAsnSerLeuSerThrIle 1430
 3919 GGGAGCAAAATCCGCGCGCGCGGTGCTGATTCAGCGCATT..... 3957
 1431 AlaAspLysThr.....GlyIleSerAspGln 1440
 3958 .CAGGCACCATACCGCGCGGTTTGGC.....GGATTGCGATCGAAGC 4000

```
1440 rlysalatylsleaspginglypheylysalalyansnglylysa 1457
1401 CGCAGCATCGCGGACGCGGATTTGTCGCAAAAGCGATTCAC 4050
1457 snllevalthrglyalagluhnglyhsthrhlaasnaalspyleysal 1473
4051 GAAAGCTCAAT 4062
1474 Thrhisalasp 1477

seqname: SwissProt_40:OMPA_RICCN

seq_documentation_block:
ID OMPA_RICCN STANDARD: PRT: 2021 AA
AC 052657; P95591; P95592; P95594; Q52667; Q52668; Q52669;
AC 052670; Q52674;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface
DE antigen) (rompa) (romp A).
GN OMPA OR RCI1273.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=Malish 7;
RC MEDLINE=94171067; PubMed=8125327;
RX Crocquet-Valdes P.A., Weiss K., Walker D.H.;
RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia
RT conorii (Malish 7 strain).";
RL Gene 140:115-119(1994).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=Malish 7;
RC MEDLINE=21442074; PubMed=11557893;
RX Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-F., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RT Science 293:2093-2098(2001).
RN [3]
RN SEQUENCE OF 8-204 FROM N.A.
RP STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RC MEDLINE=97015921; PubMed=8862558;
RX Roux V., Fournier P.E., Raoult D.;
RT "Differentiation of spotted fever group rickettsiae by sequencing and
RT analysis of restriction fragment length polymorphism of PCR-amplified
RT DNA of the gene encoding the protein rompa.";
RL J. Clin. Microbiol. 34:2058-2065(1996).
RN [4]
RN SEQUENCE OF 953-1012 FROM N.A.
RP STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RC Raoult D., Fournier P.E., Roux V.;
RT "Phylogenetic analysis of spotted fever group rickettsiae by study
RT of the outer surface protein rompa.";
RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
CC - FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).
CC - SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC - PPM: GLYCOSYLATED (BY SIMILARITY).
CC - SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL: U01028; AAI17405.1; -
DR EMBL: AE008674; AAL03811.1; -
DR EMBL: U43794; AAB49549.1; -
DR EMBL: U43798; AAB49550.1; -
DR EMBL: U43806; AAB49551.1; -
DR EMBL: U45244; AAB49566.1; -
DR EMBL: U46918; AAB49663.1; -
DR EMBL: U83440; AAC35176.1; -
DR EMBL: U83443; AAC35179.1; -
DR EMBL: U83448; AAC35184.1; -
DR EMBL: U83453; AAC35189.1; -
DR InterPro: IPR003858; rompa_OMP.
DR Pfam: PF02708; rompa_OMP; 1.
KW Antigen; Repeat; Signal; Cell wall; s-layer; Glycoprotein;
KW Complete proteome.
FT SIGNAL 1 38
FT CHAIN 39 2021
FT DOMAIN 238 946
FT DOMAIN 1424 1528
FT VARIANT 60 60
FT VARIANT 76 76
FT VARIANT 86 137
FT VARIANT 126 133
FT VARIANT 953 954
FT VARIANT 1245 1245
FT VARIANT 1308 1308
FT VARIANT 1877 1877
FT VARIANT 10 10
FT CONFLICT 92 126
FT CONFLICT 126 137
FT CONFLICT 137 157
FT CONFLICT 157 137
FT CONFLICT 368 368
FT CONFLICT 374 388
FT CONFLICT 640 640
FT CONFLICT 669 669
FT CONFLICT 793 793
FT CONFLICT 803 804
FT CONFLICT 809 823
FT CONFLICT 898 898
FT CONFLICT 908 908
FT CONFLICT 985 985
FT CONFLICT 1009 1009
FT CONFLICT 1013 1013
FT CONFLICT 1182 1182
FT CONFLICT 1314 1314
FT CONFLICT 1451 1451
FT CONFLICT 1624 1624
FT CONFLICT 1628 1628
FT CONFLICT 1872 1872
FT CONFLICT 1875 1875
FT CONFLICT 1879 1879
FT CONFLICT 1936 1936
FT CONFLICT 1965 1970
FT CONFLICT 1997 1997
SQ SEQUENCE 2021 AA; 203328 MW; 327FC42D7CB24668 CRC64;

Alignment_scores:
Quality: 231.00 Length: 1576
Ratio: 0.326 Gaps: 76
Percent Similarity: 44.924 Percent Identity: 19.670

Alignment_block:
us-09-303-518d-649 x OMPA_RICCN
Align seg 1/1 to: OMPA_RICCN from: 1 to: 2021
152 ACCATACATCGGACGCGGATTTGTCGCAAAAGCGAAGTTGTCAGTCGGG 201
```

[illegible][illegible]

```
1828 .....CTTGTTACAGCCCGCCGACGAGACCGCCCTGCTGC 1867
      : : : : : : : : : : : : : : : : : : : : : : : :
187 palagInleuInleuGlnAsnSerSergIysAsnSprgThrIleThrL 1204
      : : : : : : : : : : : : : : : : : : : : : : : :
1866 TTTCGGGGGAGAACATTTAAACGGCAAC..... 1896
      : : : : : : : : : : : : : : : : : : : : : : : :
1204 eu.....GlyAlaAsnIleAsnProAsnAsnAsnAsnGlyIleVal 1218
      : : : : : : : : : : : : : : : : : : : : : : : :
1897 .....ATCACGCAACAAACGGCA 1916
      : : : : : : : : : : : : : : : : : : : : : : : :
1219 ILeuAsnSerValThrAlaGlyIysIysLeuThrIleAlaGlyGlyIly 1235
      : : : : : : : : : : : : : : : : : : : : : : : :
1917 ACTGTT.....TTCAGCGGCA 1935
      : : : : : : : : : : : : : : : : : : : : : : : :
1235 sthrPheGlyGlyAlaHisLysLeuGlnAspIleValPheIysGlyIug 1252
      : : : : : : : : : : : : : : : : : : : : : : : :
1936 .....CCACACGCGACGCTACAAATCATTAAAC 1965
      : : : : : : : : : : : : : : : : : : : : : : : :
1252 LysPheGlyThrAlaGlyThrPheAsnThrAsnIleValLeu 1268
      : : : : : : : : : : : : : : : : : : : : : : : :
1966 GACCATTTGTCGCAAAAGAG...GGCATTCCTCGGGGAAATGCTGTG 2012
      : : : : : : : : : : : : : : : : : : : : : : : :
1269 AspIleThrGlyIleuGlnLeuGlyAlaThrThrAlaAsnValValLe 1285
      : : : : : : : : : : : : : : : : : : : : : : : :
2013 GGACACGACGTGATCACCACGACATTTAAACGGAA..... 2049
      : : : : : : : : : : : : : : : : : : : : : : : :
1285 upheliysAspAlaValAlaGlnLeuThrGlnThrGlyAsnIleGlyPheL 1302
      : : : : : : : : : : : : : : : : : : : : : : : :
2050 ..AACTTCCAAATTTAAAGCGGACGCGGTTCGCGCAATGT... 2094
      : : : : : : : : : : : : : : : : : : : : : : : :
1302 euAspPheAsnAlaLysAsnGlyThrValThrLeuAsnAsnValAsn 1318
      : : : : : : : : : : : : : : : : : : : : : : : :
2095 .....GCCAAAGTAAAGCGCATTTGACATTCAGCCCAAC 2138
      : : : : : : : : : : : : : : : : : : : : : : : :
1319 ValAlaGlyThrValLysAsnThrGlyLysAsnGlyThrLeuI 1335
      : : : : : : : : : : : : : : : : : : : : : : : :
2139 AGTTTGGTGTGCGACCGCATCAAGCCACAAATCTGTACGCTGG 2188
      : : : : : : : : : : : : : : : : : : : : : : : :
1335 eValIleuGlyAlaSerAsn.....LeuAsnThrValAla 1346
      : : : : : : : : : : : : : : : : : : : : : : : :
2189 ACTGACGCGGTGACAAATGTGTGCAAAACCAATACCGACGATAA 2238
      : : : : : : : : : : : : : : : : : : : : : : : :
1346 snGlyIleAlaMetLeuLysValAlaGlyIysValThrIleAlaLys 1362
      : : : : : : : : : : : : : : : : : : : : : : : :
2239 GTGATTCCTTCATTGACTAGACCGACATCAGCGCATCTGCATCTGC 2288
      : : : : : : : : : : : : : : : : : : : : : : : :
1363 .....GlyIysAsnValLysIleG 1369
      : : : : : : : : : : : : : : : : : : : : : : : :
2289 CGAT.....CAGCTCATTTAA 2305
      : : : : : : : : : : : : : : : : : : : : : : : :
1369 yGluIleGlnGlyThrGlyThrAsnThrLeuThrLeuProAlaHisPheL 1386
      : : : : : : : : : : : : : : : : : : : : : : : :
2306 ACTTCACAGGCGTTCACACTCAAGGC.....AAT 2337
      : : : : : : : : : : : : : : : : : : : : : : : :
1386 yLeuThrGlySerIleAsnLysThrGlyGlnAlaLeuLysLeuAsn 1402
      : : : : : : : : : : : : : : : : : : : : : : : :
2338 CTTAGTCAAAATGGCGATACAGCTTATACAGCCACCAAGCGACCA 2387
      : : : : : : : : : : : : : : : : : : : : : : : :
1403 PheMetAsnGlyLysValSerValSerGlyValAlaGlyThrAlaAlaAsnSe 1419
      : : : : : : : : : : : : : : : : : : : : : : : :
2388 AAACGCAACCTTACGCTGTCGCG.....A 2413
      : : : : : : : : : : : : : : : : : : : : : : : :
1419 rValGlyAspIleThrThrAlaGlyAlaThrSerPheAlaSerSerValAla 1436
      : : : : : : : : : : : : : : : : : : : : : : : :
2414 ATGCCCAAGCAATTTAAATCAAGCCACATTTAAAGCGCAACATCGCT 2463
      : : : : : : : : : : : : : : : : : : : : : : : :
1436 snAlaLysGlyThr.....AlaThrLeuGlyGlyThrThrSerPhe 1449
      : : : : : : : : : : : : : : : : : : : : : : : :
2464 TCGGCAATGCTTCATTATCTAAGCGACCGCGTACAAAC...GG 2510
      : : : : : : : : : : : : : : : : : : : : : : : :
1450 Ala.....HisThrPheThrAsnThrG 1457
```

```
2511 CAGTCTGACCTTTCCGCAACGCTAAGCAACGATAGCATTCGCGAC 2560
      : : : : : : : : : : : : : : : : : : : : : : : :
1457 yAlaValThrLeu.....AlaLysGlySerIleThrSerPheAla 1471
      : : : : : : : : : : : : : : : : : : : : : : : :
2561 TCACAGGTAATGCTCC.....CTACCGATAAGCGATGATTCATTT 2604
      : : : : : : : : : : : : : : : : : : : : : : : :
1471 yAsnValThrAlaThrSerPheValAlaAsnSerAlaThrIleAsnPhe 1487
      : : : : : : : : : : : : : : : : : : : : : : : :
2605 GAACGAC...CGCTTACCGGACAAATCAGCGCGCGCAAGGATACGCG 2651
      : : : : : : : : : : : : : : : : : : : : : : : :
1488 GlyAsnSerLeuAlaPheAsnSerAsnIleThrGly..... 1499
      : : : : : : : : : : : : : : : : : : : : : : : :
2652 ATTACACTTAAAGACAGCATGAGCTGCTCCATCAGGACGGAATTAG 2701
      : : : : : : : : : : : : : : : : : : : : : : : :
1500 .....SergIlyThrIleu. 1504
      : : : : : : : : : : : : : : : : : : : : : : : :
2702 GCAATTTAAACCTTGACACCGCACCATTTACATCATTCCGCTATCCG 2751
      : : : : : : : : : : : : : : : : : : : : : : : :
1505 ..ThrLeuGlyAlaAsnGlnValThrTyThr..... 1514
      : : : : : : : : : : : : : : : : : : : : : : : :
2752 CACGATCGCGGCGGCGCAACCGGACGTCGACAGATGCGCGCGCG 2801
      : : : : : : : : : : : : : : : : : : : : : : : :
1515 .....GlyThrGlySerPheThrAspThr..... 1522
      : : : : : : : : : : : : : : : : : : : : : : : :
2802 CGGTTCGCGCGTTCGCGCGCTCCCTATTCGCTTACACCGCAACTT 2851
      : : : : : : : : : : : : : : : : : : : : : : : :
1523 .....LeuThrLeuAsnThrThrPheAsnGlyAla 1532
      : : : : : : : : : : : : : : : : : : : : : : : :
2852 CGGTATATCCGTTTCACACGCTGACGCTAACGCAAA...TTGAC 2898
      : : : : : : : : : : : : : : : : : : : : : : : :
1532 IaAlaLysSerGlyLysAlaIleLeuIleLysSerGlySerThrLeuAsp 1548
      : : : : : : : : : : : : : : : : : : : : : : : :
2899 GGTGACGGAACATTCGCTTATGTCGACCTTCGCTGACCGACGAGA 2948
      : : : : : : : : : : : : : : : : : : : : : : : :
1549 LeuSerGlyValSerAsnLeuAlaLeuValAlaThrAlaThrAsnPheAs 1565
      : : : : : : : : : : : : : : : : : : : : : : : :
2949 CAATTTGACCTGCGGAAAGTTCGACGACCATTCACCTTGCGGTCA 2998
      : : : : : : : : : : : : : : : : : : : : : : : :
1565 pMetAsnAsnIleSerProAsnThrPheLysThrValIleSerAlaGluT 1582
      : : : : : : : : : : : : : : : : : : : : : : : :
2999 ACAATACCGGCAACGACCTGCAAGCTCGAACATTCGACGATGAGAA 3048
      : : : : : : : : : : : : : : : : : : : : : : : :
1582 hrAlaGlyLysLeuLysProThrProLysGlnValLysIleThrIle 1598
      : : : : : : : : : : : : : : : : : : : : : : : :
3049 GGAAGAAGCAACAAACCGCTGCTC.....GAACCTTAATTTAC 3089
      : : : : : : : : : : : : : : : : : : : : : : : :
1599 AsnAsnAspAsnArgPheValAspPheThrPheAspAlaSerThrLeuTh 1615
      : : : : : : : : : : : : : : : : : : : : : : : :
3090 CCTGCAAAACGACACGTCGATGCGCGGCGGTTCACCACTCATCC 3139
      : : : : : : : : : : : : : : : : : : : : : : : :
1615 rLeuPheAlaGluAspIleAlaIleAlaGlyVal..... 1625
      : : : : : : : : : : : : : : : : : : : : : : : :
3140 GCAAAGACGCGGATCCGCTGCAATAATCCGTCAAAGAACAGAGCTT 3189
      : : : : : : : : : : : : : : : : : : : : : : : :
1626 ..IleAspGluAspPheAlaProGlyIlyProLeu..... 1636
      : : : : : : : : : : : : : : : : : : : : : : : :
3190 TCCGCAACATCCGGAAGCGAGCAACCAAAACAGCGGCAAAAGACA 3239
      : : : : : : : : : : : : : : : : : : : : : : : :
1637 .....AlaAsnIleProAs 1641
      : : : : : : : : : : : : : : : : : : : : : : : :
3240 CGCGCAACCTTGACGCGCTGATTCGCGCGCGCGGATCCGCTGCAAA 3289
      : : : : : : : : : : : : : : : : : : : : : : : :
1641 nAlaAlaAsnIleLysLysSerLeuGluLeuMetGluAspAlaProAsnG 1658
      : : : : : : : : : : : : : : : : : : : : : : : :
3290 AGACGAAAGCGTTCCGACACCGCGCGGCGGAGGCGGGAATAATGTC 3339
      : : : : : : : : : : : : : : : : : : : : : : : :
1658 LysSerAsp.....AlaArgGlnAla...PheAsnAsnPhe 1668
      : : : : : : : : : : : : : : : : : : : : : : : :
3340 GGCATTATTCAGCGAGAGAGAAAGAAACGGGTGACGCGGATTAAGA 3389
      : : : : : : : : : : : : : : : : : : : : : : : :
1669 GlyLeuMetThrProLeuGlnGlu.....AlaAspAlaThr 1680
      : : : : : : : : : : : : : : : : : : : : : : : :
3390 CACCGCTTGGCAAAACAGCGCAAGCGAAACCGCGCGCTACACGCG 3439
```

```

1680  |||
1680  rThHisLeu..... 1683
3440  CCTTCCCGCGCGCGCGCGCGCGCGGATTCGCCAATCGCAACCC 3489
1684  .....MetGlnAsp 1686
3490  CAACCCGAGCCCAACCGCGCGCGCGCGCGGATTCAGCCCTTATGCCAATAG 3539
1687  ValValAlaProSer.....AspThrIleAlaValAlaAsnAsnGly 1700
3540  CGGTGAGTGAATTTCCGCCAGCGCTCAACAGCGTTTCCGCTACAG 3589
1700  nValValAla.....SerAsnIleSerSerAsnIleThrAlaLeuAsnA 1715
3590  ACGAATTAGCCCGCTATTGGC..... 3612
1715  laArgMetAspLysValGlnAlaGlyLysGlyProValSerSerGly 1731
3613  .....GAAGACCGCGCGCGCGCGCGTTGGACAGCGCG...ATCCG 3650
1732  AspGluAspMetAspAlaLysPheGlyAlaThrPheSerProPheValG 1748
3651  GGACACCAACACTACCGCTTCCGCGCGCGCGCTACCGCGCAACAA 3700
1748  yAsnAlaThrGlnLysMetCysAsnSerIleSerGlyTyrLysSerAsp 1765
3701  CGGACCTCGCGCAATATGCTATGCAAGAAACCTCGCGCGCGCGCTC 3750
1765  hThrGlyGlyThrIleGlyPheAspGlyPheValSerAspLeuVal 1781
3751  ...GGCATCCGTTTTCGACACCGCGCGCGCGAACCCTTGACGAC... 3795
1782  LeuGlyLeuAlaThrIleThrAlaAspThrAspIleLysLeuLysAsnA 1798
3796  ...GGCATCGCACTCGCGCGCGCGCTCCCGCGCGCGCGCTTTCG 3838
1798  nLysThrGlyAspLysAlaValGlnSerAsnIleThrSerLeuTyrG 1815
3839  GGCAATACGGCATCGACAGCTTCTACATCGCGATCGCGCGCGCGGT 3888
1815  LysLeuThrSerValProTyrGlnLeuPheValGlnAlaIleAlaSer 1831
3889  TTTAGAGCGCGCGCGCTTTCAGACGGCATCGAGCGCAAAATCCG... 3933
1832  TyrSerAsp.....AsnLysIleAspSerIly 1840
3934  ...CGCCGCGTGTGTCATTAAGCGCATTCAG...GCACGATACCGCGCG 3976
1840  sSerAlaGlyValIleAlaThrThrLeuGlnThrValGlyTyrGlnTh 1857
3977  GTTTCGGCGGATTCGGCATCGACCGCGCATCGCGCGCGCGCTATTTC 4026
1857  laAsnGlyLysTyrLysSerGlnSerIleThrGlnLeu..... 1870
4027  GTCCAAAAGCGGATTAACCGCTAC.....GAAAGCTCAATATCCG 4067
1871  ....MetAlaGlyTyrThrTyrMetMetSerGlnAsnIleAsnLeu... 1884
4068  CACCCCGCGCTTGCATTCACCGCTACCGCGCGCGCGCATTAAGCGCAT 4117
1885  ThrPro...LeuAlaGlyLeuThrTyr...SerThrIleLys...AspL 1898
4118  ATTCATTCAAACCGCG...CAACACATTCCTCAACCGCT 4155
1898  ySerTyrLysGlnThrGlyThrThrTyrGlnAsnLeuThrValLysGly 1914
4156  TATTGAGCGTGTCTATACCGATGCGCGTTCGGCGCAAGTCGAACAG 4205
1915  LysAsnTyrAsnThrPheAspGlyLeuGlnAlaLysValSerSerAs 1931
4206  CGTCAATAC...GCCGATTTGG 4225

```

```

1931  nIleAsnValAsnGlnIleValLeuThrProGlnLeuTyrAlaMetValA 1948
4226  CTCAGGATTTGCGCAAAACCGCGCGCGGATGCGCGCTAAACCGCA 4275
1948  sPtyrAlaPheLysAsnLysValSerAla.....IleAspAlaArg 1961
4276  ATCAAGGTTTACGCTGTCCCTCCACGCTGCGCGCGCGCAAGCGCGCA 4325
1962  LeuGlnIleMetThrAlaProLeuProThrAsnSerPheLys.....G 1976
4326  ACTGGAAGCGCGCAACACGCGCGCGCGCATC 4353
1976  nSerLysThrSerPheAspValGlyVal 1985

seq_name: SwissProt_40:OMP-RICRI
seq_documentation_block:
ID  OMP-RICRI  STANDARD;  PRT;  1654  AA.
DI  053047;
DI  30-MAY-2000 (Rel. 39, Created)
DI  30-MAY-2000 (Rel. 39, Last sequence update)
DI  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Outer membrane protein B precursor (168 kDa surface-layer protein)
DE  (Surface protein antigen) (Cell surface antigen 5) (Scas) (TompB)
DE  (Tomp B) (Contains: 120 kDa surface-exposed protein (Surface protein
DE  antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN  OMPB.
OS  Rickettsia rickettsii.
OC  Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC  Rickettsiaceae; Rickettsiinae; Rickettsia.
OX  NCBI_Taxid=783;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=R;
RX  MEDLINE=92167802; PubMed=1724278;
RA  Gilmore R.D. Jr., Cieplik W. Jr., Policastro P.F., Hackstadt T.;
RT  "The 120 kDa surface-exposed outer membrane protein (Tomp B) of Rickettsia
RT  rickettsii is encoded by an unusually long open reading frame:
RT  evidence for protein processing from a large precursor.";
RT  Mol. Microbiol. 5:2361-2370(1991).
RN  [2]
RP  SEQUENCE OF 279-1654 FROM N.A.
RC  STRAIN=R;
RX  MEDLINE=90136087; PubMed=2515418;
RA  Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT  "Cloning, expression and sequence analysis of the gene encoding the
RT  120 kDa surface-exposed protein of Rickettsia rickettsii.";
RT  Mol. Microbiol. 3:1579-1586(1989).
CC  -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC  STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC  VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC  -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC  LAYER WITH HEXAGONAL SYMMETRY.
CC  -1- SIMILARITY: BELONGS TO THE RICKETTSIAL OMPA/OMPB FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X16353; CA934403.1; -
DR  InterPro; IPR003858; TompA_TompB.
DR  Pfam; PF02708; TompA_TompB; 1.
KW  Antigen; S-layer; Cell wall.
FT  CHAIN 1 1333
FT  CHAIN 1334 1654
FT  DOMAIN 1181 1188
FT  DOMAIN 1654 1654
FT  DOMAIN 1654 1654
SO  SEQUENCE 1654 AA; 168184 MW; D7AB70EB7087618 CRC64;

```

alignment_scores: Length: 1603
 Quality: 224.50
 Ratio: 0.316
Percent Similarity: 44.292 Percent Identity: 19.152

alignment_block:
US-09-303-518D-649 x OMPB_RICRI ..

Align seg 1/1 to: OMPB_RICRI from: 1 to: 1654

```
244 GCGAATTCATGCAAGAGCCGATGATGATTTTCGTGTCGCGC 293
|||||:.....|
130 GlyThrLeuThr.....|LeThrGlyGlnGlyValThrAs 142
294 TAACGGCGTGGCGCATTTGGTGGCGATCATATATGTGACGGCGC 343
|||:.....|
142 nlaGlnAlaAlaLeThrLysAsnAlaGlnAsnValValAlaGlnPhea 159
344 ATACGGC.....GCGTATACACAGCTT 366
|||:.....|
159 snasnglyAlaAlaIleAspAsnAsnAspLeuysgIyValGlyArgile 175
367 GATTTTGGTGGGAGAGAAATCCGATCAACATCGTTTACTATAA 416
|||||:.....|
176 AspheGlyAlaProAlaSerThr.....LeuValPheAs 187
417 AATTGTGAACGAAATATTTATAAGCA..... 444
|||:.....|
187 nleuAlaAsnProThrThrGlnLysAlaProLeuIleLeuGlyAspAsn 204
445 .....GGGACTAAAGCCCATCCCTTATGGCGGC 471
|||||:.....|
204 laValIleAlaAsnGlyValAsnGlyThrLeuAsnValThrAsnGly... 219
472 GATTATCATATGCCCGCTTGCATTAATTTGTACAGATCGAAGACCTG 521
|||:.....|
220 ...PheIleGlnValSerAsnLysSerPheAlaThr.....Va 231
522 TGAATGACCACTTATGATGGCGGAAATATATC.....GATC 562
|||:.....|
231 LysAlaIleAlaAsnIleAlaAspGlyGlnGlyIleIlePheAsnThrAsp 248
563 AAAATATTAACCTGACCTGTTCTGATGTCGCGCAGCAGCAAGCATATTGG 612
|||||:.....|
248 laAsnAsn..... 250
613 CGATCTGATGAAGATGACCCATTAACCGCAAGTTCTATTCATATTGC 662
|||:.....|
251 .....Al 251
663 AAGTGGTATTTTGGCTGCTGTGGCAATACCTTTGCCAATAATGAT 712
|||:.....|
251 aAsnThrLeuAsnLeuGlnAlaGlyGlyThrThrIleAsnPheThrGlyT 268
713 CAGGTGGT...GGCAGATCACTTAGTGTAGTAAAAAATTAACATAGC 759
|||:.....|
268 hAspGlyThrGlyArgLeuValLeuLeuSer.....LysHisAla 281
760 CCATATGGTTTTCACACAGAGGCTCATTTGGCGACAGTGGCTCAC 809
|||:.....|
282 AlaAlaThrAsnPheAsnIleThrGlySerLeuGlyLysAsnLeuysgI 298
810 AATGTTTATTCATGATGCCAAAAGCAAAAGTGTATTAATGCGGTAT 859
|||:.....|
298 yAlaIleGlnPheAsnThrValAlaValAspGlyGlnLeuThrAlaAsn 315
860 TCGAAACGGGCAACCCCTATATAGAAAAGCAAGTGGCTTCACAGCTGT 909
|||:.....|
315 laGlyAlaAlaAlaAlaValIleGlyThrAsnAsnGlyAlaGlyArgAla 331
910 CGTAAAGATGGTCTATGATGAATATTTGCTGAGATACCATTCAGT 959
|||:.....|
332 .....laGlyPheValValSerVa 338
```

```
960 ATTTCAGACACACCTCAAAATGGGAAATACTCTTTTAACGAGATAT 1008
|
338 l.....AspAsnGlyLysValAlaThrIleAspGlyGlnV 350
1009 .....AATGGCAGCAAGAAAATC 1026
|||:.....|
350 aITyRAlaLysAspMetValIleGlnSerAlaAsnAlaThrGlyGlnVal 366
1027 AATGCCAAACATGAACACAAATTCCTGCTAATAGATTAACACGAGAC 1076
|||:.....|
367 AsnPheArgHis..... 370
1077 CGTTCAATGTTTAAATGTTCTTATCCGAGACAGACAGAACACTGTT 1126
|||:.....|
371 .....IleValAspVal..... 374
1127 ATCATGCTGCAGGTGCTGCACACATATACGCC..... 1161
|||:.....|
375 .....GlyAlaAspGlyThrThrAlaPheLysThrAlaAlaSerLysVal 389
1162 AGACTGAATATGAGAAATATTCCTTATTTAGCAAGGAAAA..... 1206
|||:.....|
390 ThrIleThrGlnAspSerAsnPheGlyAsnThrAspPheGlyAsnLeuAl 406
1207 .....GCGCATTTGATCTTACCGACACATC..... 1233
|||:.....|
406 aAlaGlnIleLysValProAsnAlaIleThrLeuThrGlyAsnPheThrG 423
1234 .....ATCAAGT.....GCTGAGAGATTATATTCCAAAGA 1266
|||:.....|
423 LysAlaLysAsnProGlyAsnThrAlaGlyValIleThrPheAspAla 439
1267 GATTTTACGGTTCGCTGAAATATACGAAATCTGGCAGCGCGCGCT 1316
|||:.....|
440 AsnGlyThrLeuGlnLysThrAlaSerAlaAsp.....AlaAsnVa 452
1317 TCATATTCAGTGAACAGTACCTGTTACTTGGAAATGAAAGCGGTGCA 1366
|||:.....|
452 laValAlaThrAsnAsnIleThrAlaIle...GluAlaSerGlyAlaGly 468
1367 ACGACCGCTGTCGCAAAATCGGCAAAAGCAGCTGACAGTTCAGCCAA 1416
|||:.....|
468 alValAlaLysLeuSer.....GlyThrHisAlaAlaGlnLeuArg 480
1417 GGGGAAAACCAAGGCTCGATC...AGCGTGGCGAGCGGTACAGTCATTT 1463
|||:.....|
481 LeuGlyAsnAlaGlySerIlePheLysLeuAlaAspGlyThrValIle... 496
1464 GGATCAGCAGCAGACGATTAAGCAAAAACACACCTTTAGTGAATCG 1513
|||:.....|
497 .....AsnGlyLysValAsnGlnThrAlaLeuValG 507
1514 GCTTGGTACGGCGCAGCGAGGTGACGATGATGCGCATATGATGCTT 1563
|||:.....|
507 LysGlyAlaLeuAlaAlaGlyThrIleThrLeuAspLysSerAlaThrIle 523
1564 AACCCGAGC..... 1572
|||:.....|
524 ThrLysAspLysIleGlyAsnAlaGlyAlaAlaAlaLeuGlnArgIleTh 540
1573 .....AAACTGATTT 1582
540 rLeuAlaAsnAspAlaLysLysThrLeuThrLeuGlyAlaAlaAsnIleI 557
1583 TCGGCTTTCGCGCGGAGCGTGGATTA.....AACGGCATTCGCTT 1626
|||:.....|
557 leGlyAlaGlyGlyGlyThrIleAspLeuGlnAlaAsnGlyGlyThrIle 573
1627 TCGTTCACCGTATTCAAAT..... 1647
|||:.....|
574 LysLeuThrSerThrGlnAsnAsnIleValValAspPheAspLeuAlaI 590
```


1648ACGATGACGGCGCATGATTGTCACCAACATCAAGACAAAGAAAT 1693
590 ealathraspdlnthrglyvalalaspalasersetleuthrasnialag 607
1694 CCACCGTACCATTTACAGCAATTAAGATATGCTCAACACCGCAATTAAC 1743
607 lnthleuthrthllesnnglylys.....lleglythrilleglyalaasn 621
1744 AAC.....ACCTGATAGCAAAAAAAGAAATGCCCTA 1775
622 Asnlystrleuglylnpheasnilleglysetserlystrhvalleuse 638
1776 CAACGGTGGTTGGCGAGAAAGAT..... 1800
638 rnsnglyasnvalalalleasnnglnleuvalilleglyasnaspolyalav 655
1801ACGACCAAAACGAC 1815
655 alglnphealathraspthrlytleuilethrarqthrthrasnialala 671
1816 GGGCGGCTCAACCTTGTTCACGAGCCCGCAGAGACCGACCCCTGCT 1865
672 glylnglylysallepheasnprovalalasnasnnglythr..Th 687
1866 GCTTCGCGCGGACCAATTTAAAGCGACATCACCAACAAACGCA 1915
687 rleuhalalaglythrlnleugly.....Serlathrasn... 699
1916 AACGTTTTCAGCGGACCAACACCGCAACGCTTCAATCATTTAAC 1965
699 699
1966 GACCATGTGCGCAAAAAGAGGATTCCTCGCGGCAATCGTGCGGA 2015
700Proleualagluile..... 704
2016 CAACGACTGATCAACCGCACATTAAAGCGAAACTTCCAAATTTAAAG 2065
705 710
2066 GCGGACAGGCGGTGTTCCCGCAATGTGCCAAAGTGAAGGCGATGCG 2115
710 lyvalasnvalaspthrvalleuasnvalaglylu..... 721
2116 CATTTGACATCACGCGCAAGAGTTTGTGTCGACCGCATCAAG 2165
722Glyvalasnleuthralath 728
2166 CCACACATCTGTACAGTTCGACCTG.....ACGG 2197
728 rasnillethrthraspallasnvalaglyserphevalpheasnialag 745
2198 GTCTGCAAAATGTGTGCAAAAAAACCAT..... 2226
745 lyglythrlnsnllevalserglythrvalglyglylnglnlysnlyls 761
2227ACGACGATTAAGTGAATGCTTCATTTGACTAA 2258
762 pheasnthrvalalaleuglnunsnlythrthralyspheeluglyls 778
2259 GACCGACATCAGCGCAATGTGATCTGCGATCAGCGTCACTCAATTAATC 2308
778 nalathrpheasnnglyasnthrthrllealalalasnserthrleuglnl 795
2309 TCACAGGGCTTGCACACTCAACGCGCAATCTTACTGCAATAGCGATACA 2358
795 leglyglyasnlythrthralaspysvalalaseralaspoly..Thgl 811
2359 CATTATACAGTCAGCCACACAGCCCAACAAAGCAACCTTACCTCGT 2408
811 yllevalaglnphevalasnthrnglyprollethrvalthrleuasnlysg 828
2409 GGGCAATGCCCAAGACATTTAAT.CAAGCCATTAAAGCGCAACACA 2457

828 lnhalalaprovalasnvalaleuylsglnllethrvalserglyprogly 844
2458 TCGGCTTCGGGCAATGCTTCAATTTATTCAGGACCCAC..... 2496
845 Asnvalalalleasnnglylleglyasnalaglysnhlshlsiglyalava 861
2497GCCCTCAAAAGCGAGCTG..... 2517
861 lthraspthrthlealaphelunsnserleucllyalavalpheel 878
2518ACGCTTCGCGCAC 2532
878 euproarglylepropheasnaspalaglyasnthrmeteproleuthr 894
2533 GCTAAGCAACGTAAGCCATTTCGCACTCAACGCGTAATGCTCCAGC 2582
895 llelysetthrvalglyasnlystrhalysgly...pheaspvalpr 910
2583 CGATAAGCAGTATTCATTGTAAGAGCGCGCTTTACCGGACAAATCA 2632
910 oservalvalalleuglyvalaspservalillealaspolyglvalaI 927
2633 GCGGC.....GCCAAGATACGCA 2652
927 leglyaspglasnasnillevalglyleuglyleuglyseraspasnngly 943
2653 TTACACTTAAGACAGCGAATGGAGCTGCGCGTACAGCACGGAATAGG 2702
944 llellevalasnialathr.....Thleuthrlylaly...llese 956
2703 CAATTTAACCTTACACAGCCACCATTAATCAATTCGCTATCGCC 2752
956 rthrleuasnasnnglnnglythrvalthrleuserglylyval...P 972
2753 ACGATGCGCGCGGCGCAACCGCGTGCAGACAGATGCGCGCGCC 2802
972 roasnthrproglythrvaltyrleuglythrlylleglyalaser 988
2803 CGTTCGCGCGCTGCGCGCTCCCTAATTATCCGTTACACCGCACTTC 2852
989 lyspheelysgl.....Valthrpherthrlns 998
2853 GGTGAATCCCGTTTCAACGCTGACGGTAAC..... 2886
998 ptyrasnasnleuglyasnillelealathrasnalathrleasnasp 1015
2886 2886
1015 lyvalthrvalthrthrglyglyllealaglylleglypheasnlyls 1031
2887GCCAATTTGAAGGTGACGAGAACATTCGCTTATGTCGA 2927
1032 llethrleuglyservalasnnglyasnvalalargphealasp 1048
2928 ACTCTTCGCGCTACCGGACAGCAATTAAGCTGCGGAAGTTCGAG 2977
1048 ylleleuserasnserthrsermetlleglythrthryslalasnng 1065
2978 GCACCTACACCTTGGCG.....GTCAACATATCCGGAACGAA 3015
1065 lythrvalthrlyrleuglyasnalphevalaglyasnilleglyaspser 1081
3016CCTGAAGCCTGCAACATTAAGCTAGTGAAGGAAGAACAA 3059
1082 asprthrprovalalaservalargpether.....glyseraspse 1095
3060 CAACCGCTGTCGCAACCTTAATTTACCCCTGCAACAAACAGCAGTCG 3109
1095 r.....glyalaglyleugnglyasnilletyrserglnvalillea 1109
3110 ATGCGGCGCGTGGCTTACCACTCAACGCAAGCGCGAGTTCGCC 3159

```

1109 ssphegltthrtyrasnleuglyllevalasnser..... 1120
3160 CTGCATATCCGGTCAGAAAGACAGACGCTTCCGACAACTCGCGGACG 3209
1121 .....AsnIlelle.....LeuIlelylgl 1127
3210 AGAAGCCAAAACAGCGCGGAAAAGACAGCGCAAGCGCTTACGCGCG 3259
1127 ythrtthralleasnnglyllysleaspleuvalthrasn...Thrlent 1143
3260 TGATTGCGCGCGCGCGATGCGTCGAAAAGACAGACGCTTCCGCGAA 3309
1143 hrphalsergltythserthrtrpelyasnasntrserllegluThr 1159
3310 CCGCGCGCGCGCGCGCGGAAAATGTCGCG...ATFATGAGCGCGGA 3356
1160 ThrleuthrleuAlaasnngly...AsnIleIleIleIleValIleleugl 1175
3357 GGAGAGAAAACAGCGCGCGATMAAGACAGCGCTTGGCGAAC 3406
1175 ucllyalaglnvalasntrthrtthrtthrtthrtthrtthrtthrt 1192
3407 AGCGCGAGCGGAAACCGCG...CCGCTACCGCGCGCTTCCCGCGCGC 3453
1192 InaspsnAlaasnAlaasnPheserGlyThrGlnThrTyThrleuile 1208
3454 CGCGCGCGCGCGCGCGGATTGCGGACAGTCGAAACCGCGCGCGCA 3503
1209 GlnGlyIyAlaArphasnnglyThrleu...SerProA 1222
3504 ACCGCGCGCGCGCGCGATGACCGCTTATGCGCAATACGCTTGGTGA 3553
1222 snpheAlaValThrGlyserAsnAlarph...Valasn 1233
3554 TTTCGCGCGCGCTCAACAGCGTTTTCGCGCTAC.....AGGACGAA 3594
1234 TyrtserleuilearGlyAlaAlaasnnglnaspyrValIleThrArGthas 1250
3595 TTAGCGCGCGCTATTTGGCGGAAAGCC.....GCCGCAAGCGCGCTTGG 3638
1250 nasAlaGlnasnValAlaThrAsnspIleAlaasnserPropheGly 1267
3639 AAGCGCGCGCGCGGACACCAACACTACGCTTTCGCAAGTTTCGCGCG 3688
1267 lYAlaProGlyAlaAspGlnasnValThr..... 1277
3689 ACCGCGCAAAACCGACCTCGCGCAATCGCTTTCGCAAGAAACCTCGCG 3738
1277 .....PheVala 1280
3789 CGAGCGCGCATCGGCACTCGGACGCGTTGCCACGCGCGCTTTCG 3838
1280 snAlaThrAsnThrAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1296
3839 GGCATATCGCGCATCGGCAAGTTTCAACATCGGACGCGCGCGCGGT 3888
1297 AlaAsnserAlaAsnPheserGlyAlaAlaAlaAlaAlaAlaAlaAla 1313
3889 TTTAGCGCGCGCGCTTTCAGACGCGATCGGAGGCAAAATCGCGCGCG 3938
1313 alIerhrasnValAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1330
3939 CGTGTGCGCTTACGCGCATCGGCAAGTTTACCGCGCGCTTTCGCGGAT 3988
1330 lYAsnArGleuGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1346
3989 TCGCGATCGACCGCATCGGCGGACGCGCTTATTCGTCGCAAAACG 4038
1347 MetAlaGly...ProGlnAlaGlyAlaAlaAlaAlaAlaAlaAlaGly 1362

```

```

4039 GATTACCGCTACGAAACGTCATATCGCACCGCGCGCTTCATTC.. 4086
1363 AspeluAlaIleaspsnValAlaIlyrGlyIleThrAlaLysProphety 1379
4087 .....AACCGCTACCGCGCGCGCATTTAGGACGATTTTCATTC 4126
1379 rthAspAlaAlaIleGlnserLysLysGlyLeuAlaGlyAlaAla 1396
4127 AACCGCGG.....CACACAT 4143
1396 ystrthrtthGlyAlaValIleGlyLeuAspThrleuAlaasnAspAla 1412
4144 TCCATACAGCGCTTATTTAGCGCTGCTTATACGATCGCGCTTGGCGAA 4193
1413 MetIleGlyAlaAlaIleGlyIleThrLysThrAsp..... 1424
4194 AGTCGAAACCGCGCTCAATACCGCGCTTGGCTCAGATTTGGCGAAA 4243
1425 .....IleLysHsIleGlnAspTyrlLysLys 1432
4244 CCGCGAGTCGGGAATGGGCGTAAACCGCGAATCAAGTTTCACGCTG 4293
1433 .....GlyAspLysThrAspAlaasnnglyPheserPhe 1443
4294 TCCCTCCACGCTGCGCGCGCAAGCGCGCACTGGAAGCGCAACAG 4343
1444 SerleuTyrlGlyAlaGlnLeuValLysAsnPheserAlaGlnGly 1460
4344 CCGG 4347
1460 rAla 1461
seq_name: swissprot.40:PHAB-BORPE
seq_documentation_block:
ID PHAB BORPE STANDARD: PRT: 3591 AA.
AC P12255;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Filamentous hemagglutinin.
DE PHAB.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP MEDLINE=90355839; PubMed=2388559;
RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Genetic characterization of Bordetella pertussis filamentous
RT haemagglutinin: a protein processed from an unusually large
RT precursor."
RT Mol. Microbiol. 4:787-800(1990).
RN [2]
RP MEDLINE=89202384; PubMed=2539596;
RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Filamentous hemagglutinin of Bordetella pertussis: nucleotide
RT sequence and crucial role in adherence."
RT Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).
CC -!- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND
CC INFECTION.
CC -!- SUBCELLULAR LOCATION: SURFACE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```



```
1682 AAGCAAGATCCACCGCTTACATTAAGCAATTAAGATATTCTACA 1731
    ::::: |||
2696 euGlutlIeGluy.....GlySerValAspAlaIaHis 2707
    ::::: |||
1732 ACCGGCAATACACAGCTTGATAGCAAAAAAGAAATGGCTTCACGG 1781
    ::::: |||
2708 ThrAspLeuSerValAlaIaArgAspAlaIaIaIaAsp... 2723
    ::::: |||
1782 TTGGTTGGGAGAAAGATACAGCAAAAGCAAGCGGCGCTCAACCTTG 1831
    ::::: |||
2724 ....PheAlaHisIaIaGluHisIaGluSerValaIaArgIleuSerLeuG 2739
    ::::: |||
1832 TTACAGACCCCGCCAGAAAGACCGACCTGCTTCCGGGCGGCAACA 1881
    ::::: |||
2739 LyAlaIysValIaIaIa.....GlyIleuTyr 2747
    ::::: |||
1882 AATTTAAGCGCAACATACAGCAAAAGCAAGCAAGCTTTTCAGCGG 1931
    ::::: |||
2748 GluAlaGlyPheSerLeuGlySerGlyLeuGluAlaHisIaIaG 2764
    ::::: |||
1932 CAGA.....CCAAACCGCAAGCTTACATCAATTAA 1963
    ::::: |||
2764 YArgGlyMetThrAlaIaIaGluValIaIaIaGlyTyrAlaGlyAlaSerH 2781
    ::::: |||
1964 ACGACCATTTGGTCCAAAAAGAGGCAATTCCTCGCGGGAATCGTGCG 2013
    ::::: |||
2781 IsGluGlnSerSerGluThrGlu..... 2788
    ::::: |||
2014 GACAAAGCATGATCAACCGCAATTAAGCGGAAACCTCCAAATTA 2063
    ::::: |||
2789 .....LysSerTyrArgAsnAlaHisLeuAsnH 2799
    ::::: |||
2064 AGGCGGACAG.....GCGGTGG 2080
    ::::: |||
2799 yGlyGlySerValIaIaIaGlyAsnValIeAspIleGlyIaIaAspI 2816
    ::::: |||
2081 TTTCCCGCAAT..... 2091
    ::::: |||
2816 IeAsnArgAsnArgTyrGlyGlyAlaIaIaIaGlyAsnAlaGlyThrGlu 2832
    ::::: |||
2092 .....GTTGCCAAAGTGAAGGCGATGGCATTTGAG 2123
    ::::: |||
2833 GluAlaIeAsnArgMetArgAlaIaIaIaGlySerThrIaIaIaIaIa 2849
    ::::: |||
2124 CAATCAAGCCCAAGCAAGTTTGGTGGCGACCGCATCAAGCAACAA 2173
    ::::: |||
2849 rGluGlnThrSerGlnSerGlyTyrPseValIaIaIaIaIaIaIaIa 2866
    ::::: |||
2174 TCTGTACACGTGCGACATG.....ACGGGTCTGCACAAATGT 2211
    ::::: |||
2866 IAspAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 2882
    ::::: |||
2212 GTCCAAAAAACCATTACCGGAGCATTAAGTATGCTTATGACATAAGC 2261
    ::::: |||
2883 ValIaIaGlnAsnValIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 2891
    ::::: |||
2262 CGACATCAAGCGCAATGTCATCTGCGCATACGCTCATTTAAATCTCA 2311
    ::::: |||
2892 .GluIleArgIleGlyIleu...MetAlaIaGlnValaIaIaIaIaIa 2907
    ::::: |||
2312 CAGGCGTTGCCACA.....CTCAAGCGCATCTT 2340
    ::::: |||
2907 hGlnIleValIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 2923
    ::::: |||
2341 AGTGAATAGGATACAGCTTATACAGTACGCAACCGCAACCGCAAAA 2390
    ::::: |||
2924 SerAlaAspPheAspSerSerHis.....SerArgSerThrSerGlns 2938
    ::::: |||
2391 C.....GCCAACTTACGCTTCGTGGCAATGCCCAAGCA 2425
    ::::: |||
2938 nThrGlnTyrIleuGlyIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 2951
    ::::: |||
2426 CATTTATACAGCACATTAACGCAACACATCGGCTCGGCAAT... 2472
```

```
2951 hGlnIleValAspAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 2967
    ::::: |||
2473 .....GTTCAATTAATTA 2492
    ::::: |||
2968 ValSerLeuIysAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 2984
    ::::: |||
2493 CCAGCGCGCTCAAAAGCGGACGCTGACGCTTCCGCAACCGTAAAGCA 2542
    ::::: |||
2984 rPheGlySerTyrSerGlySerHisAsnPheHisIaIaIaIaIaIaIa 3001
    ::::: |||
2543 ACCTAAGCAATTCGCGCATCAACGTAATGCTCCCTA.....GCC 2583
    ::::: |||
3001 snIleGlyAlaAsnAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 3017
    ::::: |||
2584 GATAGGACATGATTCATTTTGAAGACGCGCTTACCGGCAACA..... 2628
    ::::: |||
3018 GlyMetGlyThrSerHis.....GlnIleThrAsnGlyThrGlyThrTy 3033
    ::::: |||
2629 .....ATCAGCGGCGGCAAG 2644
    ::::: |||
3033 rAlaGlyThrSerValAspAlaIaIaIaIaIaIaIaIaIaIaIaIa 3050
    ::::: |||
2645 ATACGGCATTCACCTTAAGACAGCAATGACGCTGCGTCAGGACAG 2694
    ::::: |||
3050 sPLeuAsnLeu.....SerGlySer 3056
    ::::: |||
2695 GAATTA.....GCCAATTAACCTTGAC.....AAGC 2723
    ::::: |||
3057 ArgValaIaIaGlyIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 3073
    ::::: |||
2724 CACCATTAACATCAATTCGCGCTATCCGCAAGTACGCGGAGGCGCA 2773
    ::::: |||
3073 aThrSerIysGlnAspGluArgAsnTyrAsnSerSerGlyIaIaIaIa 3090
    ::::: |||
2774 CCGGACATCGACAGATGCGCGCGCGCGCTGCGCGCGCTGCGCGCG 2823
    ::::: |||
3090 sPAlaSerAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 3101
    ::::: |||
2824 TCCCATTAATTCGCTTACCGGCAATTCGCTGAGTAATCCGTTCAAC 2873
    ::::: |||
3102 .....LeuValaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 3115
    ::::: |||
2874 GTCGACGCTTAACCGCAATTCGCTGACGCAATTCGCGCTTATGCT 2923
    ::::: |||
3115 rGlnHisAspAsnSerArgLeuThrAsnAspIaIaIaIaIaIaIaIa 3132
    ::::: |||
2924 CCGAATCTTCGCTACCGGACGCAATTCGCTGCGGAAAGTTCC 2973
    ::::: |||
3132 Ia.....SerAspGlyLeu..... 3136
    ::::: |||
2974 GAAGGCACTTACACCTTGCGGTCACATACCGGCAACGCAACGCGCA 3023
    ::::: |||
3137 ThrGlnHisValIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 3153
    ::::: |||
3024 CCTCGAACAATTCGAGTGAAGGAAAGCAACAAACCGCTGTCG 3073
    ::::: |||
3153 pLeu.....SerGlyIysGlnIleuIaIaIaIaIaIaIaIaIaIa 3164
    ::::: |||
3074 AAAACCTTAATTCACCTCGCAAAAGCAACGCTGCGGCGGCTGG 3123
    ::::: |||
3164 LyAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 3176
    ::::: |||
3124 CCGTACCAATTCATCCGCAAAAGCGGAGTTCGCTGCAATTCGCT 3173
    ::::: |||
3177 .....LysAspGlyIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 3188
    ::::: |||
3174 CAAGAACAAGAGCTTTCGCAAACTCGCAAGCA..... 3210
    ::::: |||
3188 eSerSerThrThrIleuAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 3205
    ::::: |||
3211 .....GAAGCAAAAGCAAGGCGGAAAGCAACGCGCA 3246
    ::::: |||
```

```

3205 laGlyGluAspTyrGlnAlaGlnAlaThrIleAspValGlyGln 3221
3247 ACCCTTGAC...GGCTGATGCGCGGG... 3273
3222 ThrLysAspProAlaArgLeuGlnValGlyGlyValLysGlyThrLe 3238
3274 ...CGCATGCCCTGCAAAAGACAGAAAGCCCTGGCAACGGCGCGGC 3319
3238 uAsnGlnAspAlaAlaGlnAlaThr...ValValGlnArgAnLysHisT 3254
3320 AGCGAGCGCGG...GAAATGTGGCATATGAGCGGAGAGAG 3363
3254 rPalGlyGlyGlySerGln.PheSerValAlaGlyLysSerLeuLys 3270
3364 AAAAAGSGGTGACGCGGATAAAGACACCGCTGGCAAAACAGCGGA 3413
3270 sLysAsnGlnValArgProValGlnThrProThr...ProAspValAla 3286
3414 AGCGAAGCGCGCGCTACACCGCTTCCCGCGCGCGCGCGCC 3463
3286 spLysProProSerArg...ProThrThrProProAlaSerPro 3299
3464 GCGGAGATTGCGGCACTGCAACCCCAACCGCAGCCCAACCGCAGCGC 3513
3300 GlnProIleArgAlaThrValGlnValSerSerProProValSerVa 3316
3514 GACCTGATCAGCGCTTATGCCA...ATAGCGTTTGAAGTAAAT 3554
3316 lAlaThrValGlnValValProArgProLysValGlnThrAlaGln 3333
3555 TTCGCCGACGCTCAACAGCGTTTCGCCGTACAGAGCATATGACCGCG 3604
3333 euProProArgProValAla... 3341
3605 TATTTGCCGAGACCGCGCGCAGCGCTTGGACAAAGCGCATCCGCGAC 3654
3342 ValValProValThr... 3347
3655 ACCAAGACCTACCGCTTGCAGAT...Pr 3347
3347 oProLysValGlnValAlaLysValGlnValAlaProArgProLysVal 3364
3680 ...TCCGGCGCTACCGCGCAACAAACGACCTGCGCCAAATGGTATGC 3724
3364 luthrAlaGlnProLeuProProArgProValAlaGlnLysVal... 3379
3725 AGAAAGACCTGCGCAGCGCGCGTGGCATCTCTTTCGACACAGCGG 3774
3380 ...ThrThrProAlaValGlnProGlnLeuAlaLysValGlnThrVal 3395
3775 ACCGAAACACCTTGCAGACG...G 3797
3395 nProValLysProGlnThrThrLysProLeuProLysProLeuProVal 3412
3798 CATGGCAACTCG...CACGGTTCGCCACG 3826
3412 lAlaValThrLysAlaProProProValValGlnThrAlaGlnProLeu 3428
3827 GCGCGCTTTGGGCAATACGCGATCGACGTTTCTACATCGCATCAGC 3876
3429 ProProValLysProGlnLysAlaThrProGlyProValAlaGlnVal 3445
3877 GCGGCGCGGTTTAAAGCGCGGAGCTTTCAGACGGCATCGGAGGCAA 3926
3445 yLysAlaThrVal...ThrThrVal.GlnValGln 3455
3927 AATCGCGCGCGGCTGCTGATTAAGGCAATTCAGGACATACCGCGCGC 3976
3456 SerAlaProProLysProAla... 3462
3977 GTTTCGCGGATTTCGATCGAAGACGACATCGCGCAGCGCTATTTC 4026
3462 ... 3462

```

```

4027 GTCCAAAAGCGATTACCGCTACGAAACGTCATATGCCACCCCGG 4076
3463 .....ProValAlaLysGlnProAlaProAlaProLys 3473
4077 CCTGATTCACACCGCTACCGCGCGGCGCATTAAGCAGATTATTCATCA 4126
3474 ProLysProLysProLysProLysAlaGln.ArgPro.....L 3486
4127 AACGCGCGCAACATTCATCAGCGCTATTGAGCGCTGCTTACC 4176
3486 ySProGlyLys.....Thr 3490
4177 GATGCGCTTGGGCAATGCCAGACGCGGTCAATACCGGTATTCG.. 4224
3491 ThrProLeuSerGlyArgHisValValGlnGlnGlnValGlnValLeuG 3507
4225 .....GCTCAGGATTTCGCAAAACCGCGAGTGGGAAATGGCGC... 4263
3507 nArgGlnAlaSerAspIleAsnThrLysSerLeuProGlyGlyLysL 3524
4264 .....GTAACGCGCAATCAAA 4281
3524 euProLysProValThrValLysLeuThrAspGlnValGlnLysProGln 3540
4282 GGTTCACGCTGCTCCTCCACGCTGCGCGCGCA.....GCCCGCA 4325
3541 ThrTyrThrIleAsnArgArgGlnAspLeuMetLysLeuAsnGlyLysVa 3557
4326 ACTGGAAGCGGCAACAGACGCGGCGCATCAATATGAGCTACCGC 4368
3557 lLeuSerThrLysThrThrLysGlyLeuGlnGlnThrPheArg 3571

seq_name: SwissProt_40:FIG2_YEAST

seq_documentation_block:
ID FIG2_YEAST STANDARD; PRT; 1609 AA.
AC P25653;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Factor induced gene 2.
GN FIG2 OR YCR089W OR YCR89W OR YCR1102.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NBLTaxID=4932;
RN [1]
RP
RX MEDLINE=92397594; PubMed=1523889;
RA Wilson C., Grisanti P., Frontali L.;
RT "The complete sequence of a 6146 bp fragment of Saccharomyces
RL cerevisiae chromosome III contains two new open reading frames."
RL Yeast 8:569-575(1992).
CC
CC FUNCTION: REQUIRED FOR EFFICIENT MATING.
CC -I- INDUCTION: BY MATING PHEROMONES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X59720; CAA4254.1;
DR PIR: S19504; S19504.
DR PIR: S25345; S25345.
DR SGD: S000065; FIG2.
SQ SEQUENCE 1609 AA; 16049 MW; 7D66AD7F85A7B852 CRC64;

```

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 215.00 | Length: | 1187 |
| Ratio: | 0.378 | Gaps: | 55 |
| Percent Similarity: | 47.936 | Percent Identity: | 20.388 |

alignment_block:
US-09-303-518D-649 x FIG2_YEAST . .

Align seg 1/1 to: FIG2_YEAST from: 1 to: 1609

[illegible]

2202 GACAAATTGTCTCGAAAAAACCATTTACCAGACATTAATAACTGATTCAT 2251
||||| :|||: |||:||||| :|||: |||: |
446 atthrlysthValaspGlyValIleThrGlnTyrValThrIpcySPOL 463
2252 TGACTAAGACC.....GACATGAGGGCAATGTGCATCT 2286
|||||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:
463 eauthrGlnThrLysSerGlnAlaIleGlyValSerSerSerIleSer 479
2287 GCCGATCACGCCCTCATTTAAATCTCAGACGGCTTGCCACACCTCAAGGGCA 2366
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
480 ValProGlnAlaSer...serPheSerGlySerSerIleLeuSerSerAs 495
2337 TCTTAGT.....GCMAATG 2350
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||:
495 nSerSerThrLeuAlaAlaSerSmsnValProGluSerThrAlaSerG 512
2351 GCAGATACCGTTAT..... 2364
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
512 LysSerGlnTyrGlnAspTrpSerSerSerSerLeuProLeuSerGln 528
2365ACAGTCAGCCACMACGCCACCCAAAAGCGCACTTAG 2402
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
539 ThrThrTrpValValIleAsnThrThrAsnThrGlnGlySerValThrSe 545
2403 CCTCGTGGCAATGCCACAGACAATTTAATCAAGCCACATTAACCGC. 2451
| :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||:
545 rThrThrSerProAlaTyrValSerThrAlaThrLysThrValAspLys 562
2452AACACATCG 2460
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
562 alIleThrGlnTyrValThrTrpCysProLeuThrGlnThrLysSerGln 578
2461 GCTTCGGGCATGCTTCATTTATCTAAGCCACGACGCCGATCAACAAC.. 2508
||| ||| :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||:
579 AlaIleGlyIleSerSerSerThrIleSer.....AlaThrGlnThrSe 593
2509GGCAGTCGTACGCTTCCGGGCACACGCTAAGCCAACGTAAGCC 2551
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||:
593 rLysProSerSerIleLeuThrLeuglyIleSerThrLeugInLeuSerA 610
2552 ATTCCGACATCAAGCT.....AATGTCGCCCTACCGCATTAAG 2589
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
610 spalarThrPheUsglyThrGlnThrIleSmsnThrIleuethrGln 626
2590 GCAGATATTCATTTGAAGAAGCAGCCGCTTACCGGACAAATCAGCGCGG 2639
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||:
627 SerThrSerIleThrGlnUpProThrTyPheSer.....GI 638
2640 CAAGGATACGGCATTTACACTTAAAGACAGCGGAATGAGCGTCCG.... 2685
| :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||:
638 yThrSerAsperPheTyrlLeucysthSrSerGluValAsnLeuAlaSerS 655
2686TCAGGACGAGAAATTGAGCAATTTAAMC 2712
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
655 erLeuSerSerTyrrProAsnPheserSerSerGlnGlySerThrAlaThr 671
2713 CTTCACACAGCCACCATTTACACTCAATTCGCCCTATGCGCACATGCGGG 2762
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
672 IleThrAnsSerThrValThrPheGlySerThrSerTyrylProser.. 687
2763 AGGGCGCAAAACCGGCACTGGAGAGAGATGGCCGCGCGCGCTTCGGGCC 2812
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
688ThSerValSerAsnProThrGlnAlaSerGln 699
2813 GTTCGGCGCGCTCCCTATATTCGTTACACCGCCCACTTCGCTGAATCC 2862
|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
699 iValSerSerSerValAsnSerLeuThrAspPheThrSerAsnSerThr 715
2863 CGT..... 2865
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||:
716 GluThrIleAlaValIleSerAsnIleHisLysThrSerSerAsnLysAs 732
2865 2865

```
732 pTySerLeuThrThrThrGlnLeuLysThrSerGlyLysGlnThrLeu 749
2866 ..TTCAACAGCGCTGACGGTAACGGCAAAATGACGGTACGAGGAAACATTTC 2913
749 aLeuSerThrValThrThr..ThrValAsnGlyAlaAlaThrGlnThr 765
2914 CGCTTTATGTCGAACTCTCGGCTACCGGAGACAAATTAACCTGAC 2963
765 rThrTrpCysProAlaSerSerThrLeaIaTyThrThrSerThrSer 782
2964 GG..... 2965
782 yStrLeuValLeuThrThrGlnValCysSerHisSerGlnCysThrPro 798
2966 AAAGTTCGAAAGCACTACACCTTGGCGGTCAACAAATACGGCAGCA 3015
799 ThrValIleThrSerValThrAlaThrSerThrThrIleProLeuLeu 815
3016 CCGTCAAGGCTCGAA.....CAATTCAGCGTAGTGAGAGAAAGACA 3058
815 rThrSerSerThrThrValLeuSerSerThrValSerGlnCylAlaLys 832
3059 ACAAAACCGCTTCGAAACCTTATTCACCTCGCAAAACGAAACAGTC 3108
832 snProAlaAlaSerGlu.....ValThrIle..AsnThrGlnValSer 845
3109 GATGCGCGCGGCTGCGCTTACCAACTACCGCAAGACGGCGAGTTCCG 3158
845 rAlaThrSerGlnAlaThrSerThrThrGlnValSerAlaThrSer 862
3159 CCTGCATATCCGGTCAAGAAC..... 3181
862 IaThrAlaThrAlaSerGlnSerSerThrThrSerGlnValSerThrAla 878
3182 .....AAGAGCTTCCGACAACTCGC 3204
879 SerGlnThrIleSerThrLeuGlyThrGlnAsnPhrThrThrThrGly 895
3205 AAGCGACAGCA.....AAAACAGCGGAAACAAACACACG..... 3241
895 rLeuLeuPhrProAlaLeuSerThrGlnUetIleAsnThrThrValVal 3242
3242 ..CGCAAGCGCTG.....ACGCGCTGATTCGCGCGCGCGCGCATG 3280
912 enrGlyThrLeuIleIleSerThrGln..ValCysSerHisSerLysC 928
3281 CCGTCAAGAAAG.....ACAGAAACGTTGCGGAACCGCGCGCGCA 3324
928 s..ValProThrValIleThrGlnValThrSerLysGlyThrProSer 944
3325 GCGCGGGAATGTCGGCATTTGCGAGCGGAGAGAGAAAGAAACGGGT 3374
945 AsnGlyHisSerSerGlnThrLeuGlnThrGlnAlaValGlnValThr 961
3375 GAGAGCGGATTAAGACACCGCTTGGGAAACAGCGGAGCGGAAACG 3423
961 uSerSerHisGlnThrValThrMetSerThrGlnValCysSerHisSer 978
3424 .....CGCGCGGCTTACACCGCGCTTC 3444
978 LeCysThrProThrValIleThrSerValGlnMetArgSerThrProPhe 994
3445 CCC..... 3447
995 ProTyLeuThrSerSerThrSerSerSerLeuAlaSerThrLysLys 1011
3448 ..CGGCGCGCGCGCGCGGATTCGCGCAACTGCAACCCCAACGCG 3496
1011 sSerSerLeuGlnAlaSerSerGlnMetSerThrPhrSerValSerThr 1028
3497 AGCCCAACG..... 3507
1028 InSerLeuProLeuAlaPhrThrCysSerGlnLysAlaSerThrThrSer 1044
3508 .....CAGCGCACCTGATCAGCCGTTA 3530
1045 ValSerGlnThrProSerAsnThrValLeuThrAsnThrIleMetSerSer 1061
3531 TCCCAATAGCGGTTTGAAGTAATTTCCGCGACGCTCAACAGCGTTTGG 3580
1061 rSerAsn..ValIleSerThrAsnGlnLysProSerSerThrThr...Ser 1076
3581 CCGTACAGGAGCAATTAACCGCGCTATTTCCGAGACCGCGCAACGCG 3630
1077 ProTyAsnPhrSer...SerGlyTySerLeuProSerSerSerThrThr 1092
3631 GTTGGACAAACCGCATCCGGACACCAAAACCTACCGTTCCCAAGATT 3680
1092 o...SerGlnTySerLeuSerThrAlaThrThrThrIleAsnGlyIle 1108
3681 CC.....GCCCTACCGCGCAACAAACCGACCTGCGCC 3712
1108 yStrThrValTyThrThrThrTrpCysProLeuAlaGlnLysSerThrValAla 1124
3713 .....AAATCGTATGCAAGAAACCTGCGCAGCGG 3744
1125 AlaSerSerGlnSerSerArgSerValAsnPhrPheValSerSerLys 1141
3745 CGCGTGGGATCCTGTTTTCGACAAACCGGAAACACCTTCGAGCA 3794
1141 sProSerSerSerLeuSerGlnThrSerThrGlnTyThrLeuSer...T 1157
3795 CGGATCGGCAACTCGGACAGCGCTTGGCCACG.....GCG 3829
1157 hrAlaThrThrThrThrIleSerGlyLeuLysThrValTyThrThrTrpCys 1173
3830 CCGTTTTCGGGCAATACG.....GCATCGACAGCTTTCATCGGCAC 3873
1174 ProLeuThrSerLysSerThrLeuGlyAlaThrThrGlnThrSerSer 1190
3874 AGCGGCGCGCGGCTTTCAGCAGCGCA.....G 3902
1190 rAlaLysValAlaGlyThrSerAlaSerAlaThrSerThrSerThr 1207
3903 CTTTCACAGCGCATCGGAGCAAAATCCGCGCGC..... 3937
1207 erLeuSerThrSerThrGlnSerGlnSerSerSerGlyTyLeuSerLys 1223
3938 ...CGGTGCTGATTAACGGCATTCAGCAGCAT.....AC 3969
1224 GlnValCysSerGlyThrGlnCysThrGlnAsnValProThrGlnSer 1240
3970 CGCGCGCGGTTTCG.....CGGATTCGGCATCGAACCGCACAT 4007
1240 rSerProAlaSerThrLeuAlaTySerProSerValSerThrSerSer 1257
4008 CGCGCGACCGCTATTCGTCGCAAAACGAGTATCCGATCAGCAAAAG 4057
1257 erSerSerPheSerThrThrThrAlaSerThrThrThrSerThrHisThr 1273
4058 TCAATATCCGACCCCGCGCTTCGATTCACACCGCTACCGCGCGCAT 4107
1274 SerVal.....ProLeuLeuProSerSerSerSerIleSerAla 1287
4108 AAGGAGCATTAATTCAAACCGCGGACACATTCATCCATCAGCGCTTA 4157
1287 rSerProSerSerThrSerLeuLeuSerThrSerLeuProSerProAla 1304
4158 TTTGAGCCTGTCTTATACG..... 4177
1304 heThrSerSerThrLeuProThrAlaThrAlaValSerSerThrPhr 1320
4178 .....ATGCGGCTTCGGGCAAAAGTCCGACAGCGGCAATACG 4215
1321 IleAlaSerSerLeuProLeuSerSerLysSerSerSerLeuSerSerPr 1337
```

4216 G 4216
1337 O 1337

seq_name: SwissProt_40:HLES_DROME

```

seq_documentation_block:
ID HLES_DROME STANDARD; PRT; 1077 AA.
AC Q02308;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Hairless protein.
GN H.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92387549; PubMed=1516831;
RA Bang A.G., Posakony J.W.;
RT "The Drosophila gene Hairless encodes a novel basic protein that
RT controls alternative cell fates in adult sensory organ development.";
RL Genes Dev. 6:1752-1769(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93041287; PubMed=1419850;
RA Walter D., Stumm G., Kuhn K., Preiss A.;
RT "Hairless, a Drosophila gene involved in neural development, encodes
RT a novel, serine-rich protein.";
RL Mech. Dev. 38:143-156(1992).
CC -1- FUNCTION: IS A POTENT ANTAGONIST OF NEUROGENIC GENE ACTIVITY
CC DURING SENSORY ORGAN DEVELOPMENT. THE EXPRESSION OF DISTINCT CELL
CC FATES BY THE TRICHOGEN (SHAFT) / TORMOGEN (SOCKET) SISTER CELL
CC PAIR DEPENDS ON THE LEVEL OF H ACTIVITY. A CERTAIN THRESHOLD LEVEL
CC OF H ACTIVITY IS REQUIRED, BELOW WHICH BOTH SISTER CELLS ADOPT THE
CC TORMOGEN FATE.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- TISSUE SPECIFICITY: OVARY, EMBRYOS, LARVAL AND PUPAL IMAGINAL
CC DISCS.
CC -1- SIMILARITY: CONTAINS A "PRD MOTIF".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M95192; AAA28607.1; ALT_INIT.
DR EMBL: X67239; CAA47664.1; -.
DR HSP: P04002; IWP.
DR Flybase: FBgn0001169; H.
KW Developmental protein; Nuclear protein; DNA-binding.
FT DOMAIN 115 123 IHR-RICH.
FT DOMAIN 642 648 POLY-SER.
FT DOMAIN 879 891 POLY-ALA.
FT DOMAIN 937 946 POLY-ALA.
FT DOMAIN 964 974 ALA-RICH.
FT DOMAIN 979 1008 HIS/PRO-RICH (PRD MOTIF).
FT CONFLICT 151 151 S -> A (IN REF. 2).
FT CONFLICT 702 703 OH -> L (IN REF. 2).
FT CONFLICT 891 891 A -> R (IN REF. 2).
FT CONFLICT 964 967 AAVA -> RLIP (IN REF. 2).
FT CONFLICT 974 974 MISSING (IN REF. 2).
SQ SEQUENCE 1077 AA; 111039 MW; A94B1A27579E2F1 CRC64;

```

alignment_scores:

Quality: 211.00 Length: 1206
Ratio: 0.373 Gaps: 54
Percent Similarity: 46.849 Percent Identity: 20.398

alignment_block:
US-09-303-518D-649 x HLES_DROME ..

Align seq 1/1 to: HLES_DROME from: 1 to: 1077

```

979 AATGGAAATATCTTTTACGACGATATATATGGCAGCAAGAAATCA 1028
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30 AsnSerSerHisSerSerAsnAsnAsnAsnGlySerSerSerAsn 46
1029 TGCCAAACATGAAACACAAATCTCTGCTTAATAGATTAAACACGAC 1078
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
46 nAspAsnSerAsnAspAspAlaSerSerSerAsnSerLysAsn 62
1079 TTCAATTTGTTTAATGTTCTTTATCCGACAGCAAGCAAGCACTTAT 1128
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63 CATGCTGAGGTGTGTCAACAGTTATCGACCCAGCAATATATGAGACA 1178
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1129 CATGCTGAGGTGTGTCAACAGTTATCGACCCAGCAATATATGAGACA 1178
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 .....SerAsnAsnAs 75
1179 AATATTTCTTTATTTGACGAGAAAGCGAATTGATATACACGCA 1228
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
75 nhrSerSerIleIleAlaGluAlaAlaAlaLysPheLeuLeuLysAsn 92
1229 ACATCATCAAGTCCTGAGGATTTATTTCCAGAGATTTTACGGTC 1278
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
92 LysLeuAsn..... 94
1279 TCGCCTGAATAATACGAACTTGGCAGCGCGGCTTCATATTCAGTGA 1328
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 .....GlySerSer.SerThr 100
1329 AGACATACCGTTA.....CTTGAAGATAAAGC 1357
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 eLyrProLeuProProProLeuProAlaAsnLeuSerArgThr 116
1358 GCGTGGCAACGACCGCTGCCAAATCGCAAGGCAAGCGCTGACG 1405
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 ThrProThrThrThrThrProSerSerSerSerSerThrAlaSer 133
1406 ....TTCAACCAAGGGAACCAAGGCTGATCAGC..... 1441
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 nGlyPheLeuProHisAlaLysThrProLysSerSerSerIleMetAla 150
1442 .....TGGCGACGCTACATCTTTTGATCGACGACGAG 1477
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150 lAserAlaAlaValAlaAlaSerValValValAlaThrAlaSerLysPro 166
1478 ACGATTAAGCAAAACCAAGCCTTGTGAATCGCTGTGTCAGCGGC 1527
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 ThrIleAspValLeuGlyValValLeuAspLysSerSerLeuGlyVal 183
1528 AGGCGTACGTTGCACTGATCCGATATCACTGACCCGCAAACT 1577
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 alaIthr.....GlySerLeuProThrThra 192
1578 CTATTTGCGCTTTCGGCGCG..... 1597
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192 lAlaValAlaAlaAlaAlaAlaGlyThrAlaLysIleGlyLysGlySerAsn 208
1597 ..... 1597
209 SerGlyGlySerPheAspMetLysArgThrProIleSerThrHisGlyAs 225
1598 .GACGTTTGATTAAAGGCGATTCG.....TTTGGTTCACCG 1637
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 nAsnSerTrpGlyLysTyrlGlyLysArgLeuGlnPhePheLysAspLyl 242

```



```
1638 TATTCAAATACGATG.....AGGGCGGATGA 1666
1642 yspheileleugleuAlaArgSerLysAspLysAspLysSerLyr 258
1667 TTGTCACCAATCAAGCAAAAGATCCACCGTTACCATTTACAGCAT 1716
259 ValSerValThrArgLysThrPheArgProProSerAlaAlaThrSerAl 275
1717 AAGATATTCGCTACACCGCAATTAACACCTTGATACCAAAAGA 1766
275 aThrValThrProThrSerAlaValThrAlaThr...ProLysAsp 291
1767 AATTGCTACACGTTGGTTGGGAGAAAGATACGACCAAAACGAA 1816
291 LuAsnSerThrSerLeuSerPheSer.....Asp 300
1817 GGGCGCTACACCTGTTTACGCGCGCGGAGAGACCGACCTGCTG 1866
301 AspAsnSerSerLeuSerSerProThrPheAlaArgAspAlaPro.... 315
1867 CTTCGGCGGAGACAAATTTAAACGCAACATCAGCAACAAACGCA 1916
316 ..TrpLysLysSerArg...ProArgArgLysLysSerLysLysSer 330
1917 ACTGTTTTCAGCGGACACCAACACCGGCGCTACATCATTTAAAG 1966
331 LeuPhePheHisArgProArgAsnSerThrLeuGlyArgAlaLeuLeu 347
1967 ACCATGCTCCGCAAAAGAGGCGCTTCGCGGAGAAATCGGTGGAG 2016
347 gThrAlaAlaArgLysArgArgArg...ProHisLysLysLysLysSer 363
2017 AAGCACTGG.....ATCAACCGGACATTTAAAGCGGAAAC.... 2052
364 GluAspLysLysLysLysLysLysLysLysLysLysLysLysLys 380
2053 .....TTCCAATTTAAAGCGGAGCGGCGGCTTCCCGCATTTG 2095
380 pasThrLeuLysLysLysLysLysLysLysLysLysLysLysLys 397
2096 CCAAGTGAAGGCGATGGCATTTGAGCAATCAGCCCAAGCACTTTT 2145
397 lAlaAlaAsp.....ThrThrThrAsn 404
2146 GGTGTCGACCGCATCAAGCAACATCTGTACACGTTGCGACGTGAC 2195-
405 GluLeuLysLysLysLysLysLysLysLysLysLysLysLysLys 421
2196 GGGTCTGACAAATGTGTGAAAGAAACCATTAACGAGATTAAGTAT 2245
421 uArgLeuGluLysLysLysLysLysLysLysLysLysLysLysLys 438
2246 CTTCATTTGACT.....AAG 2259
438 ysgLysLysLysLysLysLysLysLysLysLysLysLysLysLys 454
2260 ACCGACATCAGCGGCAATGCGATCTTGCGATACCGCTATTAATCT 2309
455 ValAspValGluMetLysMetLysThrSerGluAspLysLysLys 470
2310 CACAGGCTTGCC.....ACATCAACGCGCAATCTTACGTGCA 2347
470 tThrLysLysLysLysLysLysLysLysLysLysLysLysLysLys 485
2348 ATGGCGATACAGCTTATACAGTACAGCAACACCGCAACCA.... 2388
485 snGlyAspLeuLysLysLysLysLysLysLysLysLysLysLysLys 501
2389 AAGCGCAACCTTACG.....CTGTCGCGCAATCGCCCAAGC 2423
502 LysAlaLysLysLysLysLysLysLysLysLysLysLysLysLys 518
2424 AACATTTAATCAAGCAATTAAGCGCAACATCGGCTTCGCGCAATG 2473
```

```
518 aArgLeuLysLysLysLysLysLysLysLysLysLysLysLysLys 527
2474 CTTCATTTAATCTAAGGACACCGCTTACAAAGCGAGTGTACGCTT 2523
528 .....AlaValLysLysLysLysLysLysLysLysLysLysLys 536
2524 TCCGCAACGCTAAGCAAGCAAGTACGATTCGCGTACAGGTAATGT 2573
537 SerAspArgLysLysLysLysLysLysLysLysLysLysLysLys 552
2574 CTCCCTGACGAGTAAAGCAATTCATTTTGAAGACACCGCTTACCG 2623
552 lSerProProSerSerAlaThr..... 559
2624 GACAAATCAGCGGCGCAAGATACGATACCTTAAAGACAGCGAA 2673
560 .....AlaAlaLysArgLeuValGluLysLysLysLysLysLys 573
2674 TGACGCTGCGCTACAGCAAGCAATTAAGCAATTTAAACCTGCAAC 2723
574 ProArgLysArgLysLysLysLysLysLysLysLysLysLysLys 589
2724 CACCATTAACATCAATTCGCGCTATCGCAGATGCGGAGGCGGCAA 2773
590 .....AsnGlyCysValAsnAsnGlySerGlyValLysLys 602
2774 CCGGCGTGGCGACACATGCGCGCGCGCGCGCTTGGCGCGCT 2823
602 erGlyGlyVala.....GlyGlyLysArgSerArgAla 612
2824 TCCCTATTTATCCGTTTACACCGCACTTGGTAGAATCCGCTTCAAC 2873
613 LysGlyLysLysLysLysLysLysLysLysLysLysLysLysLys 629
2874 GCTGACGTAACGCAATTTGACGAGTCAAGCAATTCGCTTATGT 2923
629 uAlaProProGluLysLysLysLysLysLysLysLysLysLysLys 645
2924 CGGAACCTTGGCTACCGGACGAC.....AATTGAAGCTGGCGGAA 2967
646 SerThrSerProAlaThrLeuSerThrGlnProThr...ArgLeuAsn 662
2968 AGTTCGGAAGCACTTACCTTGGCGGTCAACATACCGGCAACGAC 3017
662 erLysSerLysLysLysLysLysLysLysLysLysLysLysLys 672
3018 TGCAGGCTCGAACATTTGACGCTAGTGAAGAAAGCAACAAACGC 3067
673 .....SerGlySerGlySerSerSerSerSerSer 683
3068 TGTCGAAACCTTAATTTACACCTTCGCAAAACGACGTCGATCGCG 3117
683 rGlyLysLysCysGlyAspHisProAla.....Ala 694
3118 GCGTGGCGTTTACCACTATCCGCAAGACGCGAGTTCGCGTCAAT 3167
694 lLeuSerAsnValHisHisProGlnHisSerMetLysGlnProSer 709
3168 TCCGCTCAAGAAAGCAAGAGCTTCCGCAACCTGCGAA..... 3206
710 SerSerSerLysProArgAlaLeuLeuThrSerProLysSerProAs 726
3207 .....GGCAGAACGCAAAAGAGGCGGCAAA 3234
726 lSerGlySerAsnGlyLysLysLysLysLysLysLysLysLysLys 743
3235 GACAAACGCGCAAGCTTACGCGCT.....GATTGCGCGCGG 3272
743 yLysArgSerProProLysSerAlaLysSerProValAspLysLys 759
3273 GCGGATGCGCTCGAAAGACAGAAAGCTTGGCAACGCGCGCGGAGG 3322
```

```

760 SerPheTyrArgAspPro..... 765
3323 CAGCGCGGAAAAATGTGGCATTTATGAGCGCGAGAGAGAAAAACGG 3372
766 .....TyrAlaGlyAlaGlyArgProSerThrS 775
3373 GTGACAGCGGATTAAGACACCGCTTGGCGAAGACCGGAGCGGAAC 3422
775 erGlySerAlaSerGlnAspLeuSerProArgSerSerProAlaSer 791
3423 CCGCGCGGCTACACCGGCTTCCCGG.....CG 3451
792 ProAlaThrThrProArgThrValProLysLysThrAlaSerIleArgAr 808
3452 CCGCGCGGCGCGCGCGGATTTGGCGCACTGCACACCGGACCGCGCC 3501
808 gGluPheAlaSerProSerAlaSerSerSerCysProSerProGlyA 825
3502 CAACCGGAGCGCGGCTGATCAGCGGTTATGCCAATAGCGGTTGAGTA 3551
825 sPArgSerAlaSerProProGlyArg..... 833
3552 ATTTCCGCGCAGCTCAACAGCT.....TTTGGCGGTACAGAGCAAT 3595
834 .....ArgHisMetGlnGlnGlnProHisLeuGlnArgSerSerProle 848
3596 TAGACCGCGTATTGGCGAAGACCGCGGCAACCGGTTGGACAGCGCG 3645
848 uHisTyrTyrMetTyrProProProProGlnValAsn..GlyAsnGlyse 864
3646 ATCCGCGACACCAACACTACCGGTTGCGAGATTTCCGCGGCTTACCGCA 3695
864 ValaGlySerProThr.....SerAlaPro..... 872
3696 ACAACCGACCTGCGCAAAATGGATGACAGAAAACCTGCGAGCGGCG 3745
873 .....ProHisSerAsnSerSerAlaAlaAlaValAlaAlaAlaAla 887
3746 GCGTGGCATCTGTTTGGCAACCGGACCGGAAACACCT..TCGAC 3792
888 AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 904
3793 GAGCGCATGCGCACTGCGGACGCG..... 3817
904 rThrLeuAlaAlaLeuAlaArgHisAsnProLeuThrMetHisTyrGlnT 921
3818 .....TTGCCCGACGCGCGCTTTGCGGCAATACGCGCA 3850
921 hrGlyAlaSerProLeuLeuSerProHisProGlnProGlySerAla 937
3851 TCGACAGGTTCTACATCGGCATCAGCGCG..... 3880
938 AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 954
3881 .....GCGCGGCTTTAG 3893
954 rHisAlaPheAlaTyrAsnGlyValGlyAlaAlaAlaAlaAlaAla 971
3894 CAGCGCGACGCTTCAGACG..... 3914
971 AlaAlaAlaAla..PheGlyGlnProAlaProSerProHisThrHisProH 987
3915 .....CATCGGAGGCAAAATCGCGCGCGCTGTCGATTCAGCGCAT 3956
987 sLeuAlaHisProHisGlnHisProHisProAlaAlaLeuThrThrHis 1004
3957 C.....AGCGCATACCGCGCGGTTTGGGCGATTCGATTCGATC 4000
1004 lSerProAlaHisLeuAlaThrProLysLeuThrAspSerSerThrAsp 1020
4001 CGGACATCGCGGCAACGCGCTATTTCGCAGAAAAAGCGATTCACGCTAC 4050
1021 GlnMetSerAla.....ThrSerSerHisArgThrAlaSerTh 1033

```

```

4051 GAAACGTCAATATCCGCCACCGCGGCTTGCATTCACCGCTACGCGC 4100
1033 rSerProSerSerSerSerAlaSerAlaSerSerSerAlaAlaThrSerg 1050
4101 GGGCA 4105
1050 LyAla 1051
seq_name: SwissProt_40:OMP_RICJA

```

```

seq_documentation_block:
ID OMP_RICJA STANDARD; PRT; 1656 AA.
AC 00653:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (166 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas5) (rOMPb)
DE (rOMP B) (Contains: 120 kDa surface-exposed protein (surface protein
DE antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide).
GN OMPB
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YH:
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein rOMP B of Rickettsia
RT japonica.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC SIMILARITY).
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAL OMPA/OMPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC
CC EMBL: AB003681; BAA20138.1;
CC InterPro: IPR003858; rompa_rOMPb.
CC Pfam: PF02708; rompa_rOMPb; 1.
CC Antigen: S-layer; Cell wall.
CC CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.
CC FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.
CC FT DOMAIN 528 533 POLY-GLY.
CC SEQUENCE 1656 AA; 168097 MW; 3132A6C9DD5999F CRC64;

```

```

alignment_scores:
Quality: 210.50 Length: 1479
Ratio: 0.305 Gaps: 73
Percent Similarity: 46.721 Percent Identity: 19.675

```

alignment_block:

```

US-09-303-518D-649 x OMPB_RICJA ..
Align seg 1/1 to: OMPB_RICJA from: 1 to: 1656

```

```

661 GCAGTGGCTATTCTTGGCTGTGGTGGCAATACCTTGCAGCAAAATGG 710
|||||

```

```
251 AlaSnaIaleuAsnIleuGlnAlaGlyThrThrIleAsnPheAsnG1 267
711 ATCAGTGTG...GCCACAGTCACTTAGTAGTAAAAAATTAACATA 757
1267 YTHrAspGlyThrGlyArgLeuValIleuSerIysAsnGlyAlaIat 284
758 GCCCATATGGTTTTCACACAGAGAGTGGCAGCAGTGGCTCA 807
284 hrAspPheAsn...ValThrGlySerLeuGlyCylAsnLeuIys 297
808 CCAATGTTATCTATGATGCCCAAAAGAGTGAATTAATGAGGCT 857
298 GlyIleIle...GluLeuSerThrValAlaIleAsnGlyC1 310
858 ATTC...CAACGGCAGCCCTATATAGAAAAAGCAATG 895
310 nleuIleAlaAsnAlaGlyProAlaAsnAlaValIleGlyThrAsnSng 327
896 GCTTCAGCTGTGCTAAAGATTGCTTATGATGAATCTTGTGCA 945
327 LAlaGlyArgAla...AlaGly 333
946 GATACCCATTGATTTCTACAGAACAGTCAAAATGGAAATCTCTT 995
334 PheValIleSerVal...AspAsnGlyLysAlaAlaIatH 345
996 TAACAGCATAAAT...AATG 1012
345 rIleAspGlyGlnValTyralAlaLysAspMetValIleGlnSerAlaAsn 362
1013 GCACAGAGAAAAATCATGATGCCAAACATGACACAAATCTCTGCTAATAGA 1062
362 laAsnGlyGlnValAsnPheArhIs... 370
1063 TTAAAAACGAAACCGTTCAATGTTTATGTTCTTTATCCGAGACAGC 1112
371 ..... 1113
1113 AAGAGAACCTGTTTATCATGCTGCTGAGTGGTGCACAGCTTATCCAGCCA 1162
380 r.....AlaPheIysThrAlaIleSerIleValAla..... 390
1163 GACGATATATGAGAAAAATATTTCTTTATTGACAGAGAAAA..... 1206
391 .....IleThrGlnAsnSerAsnPheGlyThrThrAspPheGlyAsnLeuAla 406
1207 .....GGCGAATGATCTTACCAGACATC..... 1233
407 AlaGlnValThrValProAspThrMetThrLeuThrGlyAsnPheThrG1 423
1234 .....AATCAAGGT.....GCTGAGAGATTATTTCCAAAGAG 1267
423 yAspAlaAsnAsnProGlyAsnThrAlaGlyValIleThrPheAlaAla 440
1268 ATTTTACGGCTCGCCTGAAAAATACGAAACTTGCAAGCGCGCGCTT 1317
440 sngIyThrLeuAlaSerAlaSerAlaAsp.....AlaAsnVal 452
1318 CATATCAGTGAAGACAGTACCGTTACTTGAAGTAAGCGCGCGGAAA 1367
453 AlaValThrAsnAsnIleThrAlaIle...GluAlaSerGlyValGlyVa 468
1368 CGACCGCTGTCCAAAATCGCAGAGCAGCTGCACGTTCAAGCCAAAG 1417
468 lValGlnLeuSer.....GlyThrHisThrAlaGlnLeuArgL 481
1418 GGGAAAACCAAGCTCGATC...AGCGTGGCGAGCGGTACGATCACTTTG 1464
481 eugIyAsnAlaGlySerValPheIysLeuAlaAspGlyThrValIle... 496
1465 GATCAGCAGCGAGCATTAAGCAAAAAACAGCCTTAAGTGAATCGG 1514
497 .....AsnGlyLysValAsnGlnThrValLeuValG1 507
1515 CTTCGTACGGCAGGAGGTACGGTCAACTGATGCCATATACGTCA 1564
507 yGlyValIleValAlaAlaGlyAlaIleThrIleuAspIleSerAlaThrIleT 524
1565 ACCCGCAACAACCTCTTTTCGGCTTCGCGCGAGCGGTGGATTAAAC 1614
524 hrGlyAsp.....IleGlyAsnGlyGlyGlyAlaAlaLeuGln 537
1615 GGGCATTCGCTTCGTTCCACCGTATTCAAAT...ACCGATGAAGGCGC 1661
538 SerIleThrLeuAlaAsnAspAlaThrIysThrLeuThrLeuGlyGlyAl 554
1662 GATGATGTTCACCAACAT.....CAAGCAAAAGAT 1693
554 AsnIleIleSerAlaAsnGlyGlyThrIleAsnPheGlnAlaAsnGlyG 571
1694 CCACGTTACCATTCACGCAATAAAGAT..... 1722
571 lYThrIleIysLeuThrSerThrGlnAsnAsnIleValValAspCysAsp 587
1723 .....ATTGCTACA...ACCGC...AATACACAGCTTGA 1754
588 LeuAlaIleAlaThrAspGlnThrGlyValValAspAlaSerSerLeuTh 604
1755 TAGCAAAAAAAGAAATGCTACACAGGTGTTGGCGAGAAAAAGATACGA 1804
604 rAsnAlaGlnThrLeuThrIleSerGlyThrIleGlyAlaIleGlyAla 621
1805 CCAAAAGCAAGC...GGCGGCTCAACCTGTTTACAGAGCCCGCGAGAA 1851
621 sAsnThrThrLeuGlyGlnPheAsnIle.....GlySer 632
1852 GACCGCACCTGCTGCTTCCGCGAGACAAATTTAAGCGCAACATCAC 1901
633 SerIysThrIleuAsnGlyLysAsnValAlaIleAsnGlnLeuValI1 649
1902 GCAAAACAAGCGCAACCTGTTTTCAGCGGAGACCCACAGCGCCT 1951
649 eGlyAsnAsnGlySerValGlnPhe.....AlaH 659
1952 ACAAT...CATTTAAGCAGCATGCTGCGCAAAAAGAGCGCATTCCTGC 1998
659 IsAsnThrIyIreuIleThrArgThrThrAsnAlaIaGly.....Gln 673
1999 GGGGAATGCTGTGGAGACAGCATGATCAACCGCATTTAAAGCGGA 2048
674 GlyLysIleIlePheAsn..... 679
2049 AAACCTTCAAAATTAAGGCGAGCAGCGGTGTTCCCGCATGTGCCA 2098
680 .....ProValValAsnAsnThrThrL 688
2099 AAGTGAAGCGCATTTGACATTCAGCCCAAGCAGT..... 2142
688 euAlaIleGlyThr...AsnLeuGlySerAlaIleAsnProLeuAlaGlu 703
2143 .....TTGGTGTGCGACCGCATCAAGCGACCAATCTGTACAGCTTC 2186
704 lIleAsnPheGlySerLysGlyAlaArgAlaAspThrValIleuAsnValG1 720
2187 GGATGAGCGGCTGTG.....ACAAATGTGTGCAAAAAACCATTAACG 2230
720 yGlnGlyValAsnLeuIyralaThrAsnIle.....ThrThrThra 734
2231 ACAGTAAAGTG.....ATTGCT 2247
734 sPaIaAsnValAlaGlySerPheValaPheAsnAlaGlyLysAsnIleVal 750
2248 TCATTGACTAAGACCGCATCAGCGCAAT.....GTGATGCT 2285
751 SerGlyThrValGlyGlyGlnGlnGlyAsnLysPheAsnThrValAlaIa 767
```

2286 TGGCGATCAGCTCATTTAAATCTACAGGGCTTCCACACTCAAGCA 2335
::: ::::: ||| ||||| ::::: |||||
UASPASNGlyThrThValIAspHeuGlyAsnAlaThrPheAsnGlyA 784
2336 ATCTTGTCGAATGGCGATACGCTTATACAGTCCGACCAAGCCAC 2385
||: ::::: ||| ||||| ||| |||
784 snThrThrIleAlaIAsnSerThrIleuGlnIleSerGlyAsnThrThr 800
2386 CA 2387
801 AlAspPheIleAlaSerAlaAspGlyThrGlyIleValGluPheValAs 817
2388 AAACGGCAACCTTACGCTGCGGCAATGCCCAAGCA.....A 2425
::: ||| ::::: ||| |||||
817 nThrGlyProIleAsnValThrIleuAsnGlyAlaValProValAsnA 834
2426 CATTTAATCAAGCCACATTAACGGCAACACATCGGCTTGGGCAATGT 2475
::: ||| ||| ::::: ||| ::::: |||
834 IAlAspGlyGlnIleThrValSerGlyProGlyAsnValValAlaValAsnGlu 850
2476 TCATTATATCAAGCGAC.....GCCGT 2501
|||
851 IleGlyAsnAlaGlyAsnThrHisGlyAlaMetThrAspThrIleAlaPhe 867
2502 ACAAAACGGCAGTCTG..... 2517
::: ||| ::::: ||| |||
867 eGluAsnSerThrIleuGlyAlaValIleuPheLeuProSerGlyIleProP 884
2518ACGCTTCCGGCAACGCTAAGCAACGTAAGC 2550
|||
884 HeAsnAspAlaGlyAsnThrIleProLeuThrIleuSerThrThrAlGly 900
2551 CATTCGCACTCAACGCTATGTCTCCCTACCGCATAGCAGCATATCCA 3260
::: ||| ::::: ||| |||
901 AsnGluThrAlaGluGly...PheSerValProSerValIleValSerG 916
2601 TTTTGAACGACCGCTTACCGGCAATCAGCGC..... 2637
::: ||| ||||| ::::: |||
916 yValAspSerValIleAlaAspGlyGlnValIleGlyAspGlnAsn 933
2638GGCAAGATACGGCATTAACCTTAAAGACAGC 2670
|||
933 IAlAlGlyLeuGlyLeuGlySerAspAsnGlyIleIleValAlaSnAlaThr 949
2671 GAATGACGCTGCCGTCAGCACGCAAGATAGGCAATTAACCTTGACAA 2720
|||
950ThrIleuThrAlaGly...IleGlyThrIleAsnAsnAsnG 962
2721 CGGCACCATTAACATTCGCTATTCGCCACGATGCGGCGAGGCGCC 2770
::: ||| ::::: ||| |||||
962 pGlyThrValThrIleuSerGlyGlyVal...ProAsnThrProGlyThrV 978
2771 AAACCGGACGTGCGACAGATGCCGCCGCCGCTTCCGCGCTTCGCGC 2820
|||
978 AlTyrglyLeuGlyThrGlyIleGlyAlaSerIlySphelySeln..... 992
2821 CGTTCCTATTATCCGTTACACGCCCAACTTCGCTAGAAATCCCGTTCAA 2870
|||
993ValThrPheThrThrAspTyrAsnAsnIleuGlyAs 1004
2871 CACGCTGACGTAAC..... 2886
|||
1004 nIleIleAlaThrAsnThrThrIleAsnAspGlyValThrValThrThG 1021
2886 2886
1021 IyGlyIleAlaIleAlaGlyIleAlaGlyThrAspPheAspGlyLysIle 1037
2887GGCAATGGAACGTCAGGAGAACATTCGCTTATGTCGGAAC 2930
|||
1038 ThrIleuGlySerValAsnGlyAsnAlaAsnValAlaThrPheAlaAspGlyI 1054
2931 CTTCGGCTACCGCAGCAATTCGAGCTTCCGGAAGTTCGGAAGCA 2980

```

1312 AlthraspThrSerAlaAlaIleThrAspAlaGlnLeuAspValAlaLys 1322
3832 ... GTTTGGGGAAATACGGCATCGACAGTTC.....TACAT 3866
      ::::||||| ||| ::::||||| ::::|
1329 AspIleGlnAlaIleuLeuLys...AsnArgLeuGlyAlaLeuArgTyrIle 1344a
      ::::||||| ::::| ::::||||| ::::|
3867 CGGCATACACGGCGGGCGGTTTACGAGCGGAGCGCTTCACAGCGCA 3916
      ::::||||| ::::| ::::||||| ::::|
1344 uGlyThrProGluMetValGlySerGlnIaIleThrProAlaIaLys 1361
      ::::||||| ::::| ::::||||| ::::|
3917 TC...GGAGCGAAATCCGCGCGCGCTGTCGATTACCGGATTTCAGCA 3963
      ::::||||| ::::| ::::||||| ::::|
1361 aLaLaIaGlyAspGlnIaValaAspAsnValaIaTyrGlyIleThrPala 1377
      ::::||||| ::::| ::::||||| ::::|
3964 CGA.....TACGCGCGGTTTCGCGG 3986
      ::::||||| ::::| ::::||||| ::::|
1378 LysProPheTyrThrAspAlaHisGlnSerLysGlyGlyLeuAlaG 1394
      ::::||||| ::::| ::::||||| ::::|
3987 ATTCGCG..... 3993
      ::::||||| ::::| ::::||||| ::::|
1394 YThrLysAlaLysThrThrGlyIleValIleGlyLeuAspThrLeuAla 1411
      ::::||||| ::::| ::::||||| ::::|
3994 .....ATCGAACCGCACATCGCGCGCACGCGCTATTTCGTC 4029
      ::::||||| ::::| ::::||||| ::::|
1411 snAsnAsnLeuMetIleGlyAlaIaIleGlyIleThrLysThrAspIle 1427
      ::::||||| ::::| ::::||||| ::::|
4030 CAAAAAGCGGATTCACCG.....TAGAAACGTCAAATATCGCCAC 4070
      ::::||||| ::::| ::::||||| ::::|
1428 LysHisGlnAspTyrLysLysGlyAspLysThrAspValAsn..... 1441
      ::::||||| ::::| ::::||||| ::::|
4071 CCGCGCGCTTCATTCAACCGCTACCGCGG..... 4101
      ::::||||| ::::| ::::||||| ::::|
1442 ...GlyPheSerPheSerLeuTyrGlyIaGlnGlnPheValGlnAsp 1457
      ::::||||| ::::| ::::||||| ::::|
4101 ..... 4101
1457 hepAlaGlnGlySerAlaIlePheSerLeuAsnGlnValLysAsnLys 1473
      ::::||||| ::::| ::::||||| ::::|
4101 ..... 4101
1474 SerGlnArgTyrPhePheAspAlaAsnGlyAsnMetSerLysGlnIleAl 1490
      ::::||||| ::::| ::::||||| ::::|
4102 .....GGCATTTAAGCAGATT 4117
      ::::||||| ::::| ::::||||| ::::|
1490 aLaIaGlyAsnTyrAspAsnMetThrPheGlyLysAsnLeuThrValGlyT 1507
      ::::||||| ::::| ::::||||| ::::|
4118 ATTCAATTCAACCGCGCAACACATTTCCATCAGCGCTATTGAGCGTCG 4167
      ::::||||| ::::| ::::||||| ::::|
1507 YrAspTyrAsnAlaMetGlnGlyValLeuValThrProMetAlaIaLysLeu 1523
      ::::||||| ::::| ::::||||| ::::|
4168 TCGTATACGATGCGCGCTTCGGGCAAAAGTCGAGAACCGCGTCAATATCCG 4217
      ::::||||| ::::| ::::||||| ::::|
1524 SetTyrLeuLysSerSerAspLysAsnTyrLysGluThrGlyThrThrVa 1540
      ::::||||| ::::| ::::||||| ::::|
4218 CGTATTGGCTCAGGATTTGGGCAAA...ACCGCGAGTCGGAAATGGGCG 4264
      ::::||||| ::::| ::::||||| ::::|
1540 lAlaAsnLysGlnValAsnSerLysPheSerAspArgThrAspLeuIleV 1557
      ::::||||| ::::| ::::||||| ::::|
4265 TAAAGCGCGCAAAATCAAAGTTTCACGCTGTCCCTC 4299
      ::::||||| ::::| ::::||||| ::::|
1557 aGlyAlaLysValaIaGlyIleThrMetAsnIle 1568

seq_name: SwissProt_40:HLVA_SERMA

seq_documentation_block:
ID HLVA_SERMA STANDARD; PRT; 1608 AA.
AC P15320;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Hemolysin precursor.
SH A.
OS Serratia marcescens.

```

| | |
|----|--|
| OC | Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae; |
| CC | Serratia. |
| OX | NCBI_Taxid=615; |
| RN | (1) |
| RP | SEQUENCE FROM N.A., AND SEQUENCE OF 31-40. |
| RC | STRAIN-SN8; |
| RX | MEDLINE=88257037; Pubmed=3290200; |
| RA | Poole K., Schiebel E., Braun V.; |
| RT | "Molecular characterization of the hemolysin determinant of Serratia marcescens." |
| RL | J. Bacteriol. 170:3177-3188(1988). |
| CC | -I- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY DEFINED. |
| CC | -I- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA REQUIRES SHLB FUNCTION. |
| CC | -I- SUBCELLULAR LOCATION: TO THE ERYTHROCYTE MEMBRANE. SHLA |
| CC | -I- SIMILARITY: TO P. MIRABILIS HEMOLYSIN (HPMA). |
| CC | ----- |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-slb.ch/announce/ or send an email to license@sb-slb.ch). |
| CC | ----- |
| DR | EMBL: M2618; AAA50323.1; -. |
| DR | PIR: A28182; A28182. |
| KW | Hemolysis; Toxin; Outer membrane; Signal. |
| FT | SIGNAL |
| FT | CHAIN 1 1608 |
| FT | HEMOLYSIN. |
| SQ | SEQUENCE 1608 AA; 165078 MW; D669BA76EE7DA0D51 CRC64; |

```
alignment_scores:  
    Quality: 210.00      Length: 1433  
    Ratio:   0.324       Gaps:     67  
Percent Similarity: 45.290 Percent Identity: 19.819
```

```

US-09-303-518D-049 x HLYA_SERMA ..
Align seg 1/1 to: HLYA_SERMA from: 1 to: 1608

```

```

649 TCATCATCAATATTCGAAGCGCATATTCCTGCTGCTGGTGGCAATACCTT 698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
251 SerIrrYrLeuGlySerMetGlnAlaGlyArgIleAsnIleLeuAsnTh 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
699 TGCACAAAATGGATCAGT.....GGTGACACATGCATCTAGCTA 739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
267 rAlaGln...GlySerValIlyLysLeuAlaGlySerLeuAsnAlaGly 283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
740 GTGAAAAAATTAACATACGCCCATGTGTTTTTACCACA..... 780
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
283 spGlu...LeuLysValIlyAlaIyrAspIleArgSerGlnSerArgVal 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
781 .....GGAGGCGCATTTGGCCACACTGG 803
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299 AspAspAlaSerSerAsnLysAsnGlyIlyAspAsnIyrGlnAsnIyrAr 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
804 CTCACCAATCTTTATATCTATGCATGCCCAAAAGCGTTATTAATAG 853
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
315 gGlyIlyIleIyrValAsnAspArgSerSerSerGlnThrLeuThrArgT 332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
854 GGGTATTTCGAAGCGGCACCCCTATATAGAAAAAGCATAGCTTCCAG 903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
332 hrcIleuLysGlyLysAsnIleSerLeuValAlaAspAsnHisAlaHis 348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
904 CTGGTTTCGTAAGAT..... 918
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
349 LeuThrAlaThrAspIleArgIlyGlnLyspIleThrLeuGlnGlyIly 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

918 ..... 918
365 sleuthrleuaspglynglnleuysglnthrglnghisthraspa 382
919 ..... TGGTCATGATGAATCTTTCCTGGAGATACCATTCAGTATTC 963
382 spargtripheryserttrp ..... glntryaspyalthr 393
964 TACGACCACTGCAAAATGCGAATCTTTTAACGACGATATATATGG 1013
394 ATGGLuArgGlnGlnleuGlnleuGlnleuGlnleuValalaasegl 410
1014 CACGGAATAATCAATGCCAATGACACAAATCTCTGCTAATGAT 1063
410 yserlalyseuileuileserthrglnlu ..... 419
1064 TAAAAACAGAACCTTCATTTGTTTAACTTTCTTTATCC ..... 1104
420 ..... AspyalLysleuLeuGlyAlaAsnValSerAlaAspArg 432
1105 ..... GAGCAGACAGAACCTGTTATCATGCTGCGAGTGG 1142
433 AlLeuSerValLysAlaAlaArgasp ..... ValHisLeuAlaGlyLe 447
1143 TGTC ..... ACAGTTATCGACCCAGACTGAATATGAGAAA 1180
447 uValGlnLysAspLysSerSerGlnuArgLysLysGlnArgasn ..... H 462
1181 ATATTTCTTTATTTAGACGAAGAAAGGCAATTTGATCTTACCAAC 1230
462 lstrSerSerLeuArgthrglyArg ..... TrpSerasn 473
1231 ATCAATCAAGT ..... GCTGAGATTTATTTCCAAAGGAGA 1268
474 SeraspLysSerGlnSerLeuLysAlaSerGlnLeuArgSerGlnLysl 490
1269 TTTTACGCTCTCGCTTAAATACGAATCTTGCAAGGCGCGGCTTC 1318
490 uLeuThrLeuLysAlaGlnArgAsnValSerThrGlnGlnLysValH 507
1319 ATATCAGTGAAGACATGACCTT ..... 1341
507 lAlaGlnArgAspLeuThrLleAspAlaAspAsnGlnLleGlnValGly 523
1342 ..... ACTTGAAAGTAAACGGCGTGCAACGCGCTGTCGA 1382
524 ValGlnLysThrAlaAsnAlaLysAlaValArgAspLysThrSerTr 540
1383 AATGGC ..... 1389
540 pGlyGlyLleGlyGlyLysAspAsnLysAsnSerAsnArgGlyl 557
1390 ..... AAAGCAGCTGCGACGT ..... 1407
557 lSerHisAlaSerGlnLeuThrSerGlyLysThrLeuArgLysAsnGly 573
1408 ..... CAGCAAGAGGCAAAACCAAG 1430
574 GlnGlnGlyValThrLleThrGlySerLysAlaArgGlyGlnLysGlyl 590
1431 CTGATCAAGCTGGCGGACGTAAGTATTTGATCAGACGACGAC 1479
590 yGlnValThrAlaThrHisGlyLysLeuArgLleAspAsnAlaLeuSerTr 607
1480 ..... GATTAAGCAAAACCAAGCTTTAGTGAATCGGCTGGTC 1521
607 hThrValAspLysLleAspAlaAspThrGlyThrAlaAsnLleThr 623
1522 AGCGGAGGAGGTAACGCTGAATGCAATCCGATATACGTTTC ..... 1563
624 SerSer ..... SerHisLysAlaAspAsnSerLysGlnSerSe 636
1564 ..... ACC 1567
636 rThrAlaSerGlnLeuLysSerAspThrAsnLeuThrLeuValSerHisL 653
1568 CCGACAAATCTATTTGCGCTTTCG ..... GCGGACGTTTGAT 1608
653 yAspAlaAspValLleGlySerGlnValAlaSerGlyGlyGlnLeuSer 669
1609 TTA ..... AACGGCATTCGTTTCGTTCCACCGTATTC 1643
670 ValGlnSerLysThrGlnLysLleAsnValLysAlaLleGlnArgGlnl 686
1644 AATATCCGATGAA ..... GCGGATGATTCAC ..... 1674
686 nAsnLleAspGlnGlnLysThrAlaLeuThrValAsnGlyTyAlaLysG 703
1675 ..... CACATCAA 1683
703 LuAlaGlyAspLysGlnTyArgAlaGlyLeuArgLleGlnHisThrArg 719
1684 GACAAAGATCCACCGTTACCATTTACAGCAATAAGATATGCTACAC 1733
720 AspSerGlnLysThr ..... ThrArgThrGlnAsnSerAlaSerSerLeuSe 735
1734 CGGCAATTAACAACAGCTTGATAGCAAAAGAAATGCTACACAGCT 1782
735 rGlyGlySerValLysLeuLysAlaGlnLysAspAlaThrSerGlyS 752
1783 ..... TGTTGGCGGAGAAAGATACGACCAAAACGAGCGGCTCAC 1827
752 erLysLeuValAlaAspLysGlyAspAlaSerValSerGly ..... Asn 766
1828 CTGTTTACCAAGCCCGCCAGAACAGCCACCTGCTGCTTCCGGGG 1877
767 LysValSerPheLeuAlaAlaAspLysThrAlaSerAsnThrGlnl 783
1878 AACAATTTTAAAGCGACATCCAGCAACAACAGCG ..... AAACGT 1921
783 nThrLysLleGlyGlyPheTyThrGlyLysLleAspLysLeu 799
1922 TTTTACGCGGACAGCAACAGCGGCTTACATCATTTTAAACGACAT 1971
800 ..... GlySerGlyValGlnAlaGlyLysLysAsnLysThrGln 813
1972 TGCTCGCAAAAGAGGCAATCTCGCGGAAATGTTGGGACACGA 2021
814 AlAlaGlnSerLysAlaLleThrSerGlySerAspValLysGlyAsnle 830
2022 CTGATCAACCGCACATTTAAAGCGGAAATCTTCAATTTAAAGCGGAC 2071
830 uThrLleAsnAlaArgAspLysLeuThrGln ..... GlnGlyAlaG 844
2072 AGCGGCTGGCTT ..... TCCGCAATGTTGGCAAGGAGGC 2109
844 lHisSerValGlyGlyAlaTyThrGlnLysAsnAlaAlaGlyValAsp 859
2110 GATTGCGCATTTG ..... AGCAATCACGC 2132
860 ..... HisLeuAlaAlaAlaAspThrAlaSerThrThrThrLysThr 874
2133 CCAACAGCTTTTGGTGTGCGACCGCATCAAGCGCAACATCTGATAC 2182
874 rAspValGlyValAsnLleGlyAlaAsnValAspLysSerAlaValThr 891
2183 GTTCG ..... GACTGACGCGGT 2199
891 rGrProValGlnuArgAlaValGlyLysAlaLysLysLeuAspAlaThrGly 907
2200 CTGACAAATTTGCTC ..... 2214
908 ValLleAsnAspLleGlyGlyLleGlyAlaProAsnValGlyLeuAspIl 924
2215 ..... GAAAAAACCATTAACGACGATTAAGTGA 2242

```

```
924 eGlylaInglYlSerSerGluYlSerSerSerSerGlnAlaVal 941
2243 TTGCTCTACTTAAGACCGACATCAGCGCAATGTCATCTT 2286
941 aValSerSerValGln.....AlaGlySerIleAspIleAsnAla 954
2287 .....GCCGATCAGCTCA 2300
955 LysGlyGluValAlaArgAspGlnGlyThrGlnTyrglnAlaSerLysGlyAl 971
2301 TTTAAATCTACA.....GGCTTCACACATCCAGCGCA 2335
971 aValAsnLeuThrAlaAspSerHisArgSerGlnAlaAlaAsnAlaG 988
2336 ATCTTAGTCCAAATGGCGATACAGT.....TAT 2364
988 InaSpGlnSerArgAspThrArgGlySerAlaGlyValArgValTyr 1004
2365 ACAGTCAGC.....CACACGCGCAC 2384
1005 ThrThrThrGlySerAspLeuThrValAspAlaLysGlyGlyGlyTh 1021
2385 CCAAAACGGCACTTAGCTCTGGGCAATGCCAAGCAACATTTAATC 2434
1021 rGlnArgSerAsnSerSerAla.....SerG 1030
2435 AAGCCACATTTAAACGCAACACATCGCTTCGGCAATGCTTCATTTAAT 2484
1030 InAla...ValThrGlySerIleAspAlaAlaAsnGlyIleAsnValAsn 1045
2485 CTAAACGACCGACCGCTACAAAAGCGAGCTGATCGCGCAACGC 2534
1046 ValLysLysAspAlaIleTyrGlnGlyThrAlaLeuAsnGlyGlyTrg 1062
2535 TAAAGCAACGTA..... 2547
1062 YLysThrAlaValAsnAlaGlyLysPLeuArgLeuAspGlnAlaSerA 1079
2548 .....AGCATTCGGCACTCAACGGTAATGTCCTCCATAGCGATTAAG 2589
1079 pRlyGlnSerGlnSerArgSerGlyPheAsnValLysAlaSerAlaLys 1095
2590 GCATATTCATTTGAAAGCAGCGCTT..... 2619
1096 GlyGlyPheThrAlaAspSerLysAsnPheGlyAlaGlyPheGlyGly 1112
2620 .....ACCGACAAATCA 2632
1112 YThrHisAsnGlyLysSerSerSerSerThrAlaGlnValGlyAsnIle 1129
2633 GCGGC.....GGCAAGATACGGCATTTACAC 2658
1129 eGlyGlnGlnGlnGlyAlaGlyLeuLysAlaGlyArgAspLeuThrLeu 1144
2659 TTTAAAGACAGCAATGAGCGCTGCCGTCAGCAGCAGCAATTA..... 2700
1145 .....GlnGlyThrAspValLysSerG 1152
2701 .GGCAATTTAACTTGACAAAGCCACCATTAACATTCGCGCATC 2749
1152 nGlyAspValSerLeuSerAlaGly.....AsnLysValAlaAl 1165
2750 GCCAGATGCGGAGCGGCGGCGAGTCGAGACAGATGCGCGCGC 2799
1165 eGlnAlaAlaGlnSerThrGlnThr.....Arg 1174
2800 CGCGCTTGGCGCGCTTCGCGCTTCCATTTATTCGTTACACGGCAAC 2849
1175 LysGlnSerLysLeuSerGlyAsnIleAspLeuGlnGlySerSerAs 1191
2850 TTCGGTAGAATCCCGTTCAACAGCGTCGAGGTAACAGCGC..... 2889
```

```
1191 pSerLysGluYlSerThrGlyLysLeuSerAlaGlyAlaPheAspI 1208
2890 .....AATTTGAACGTCAGGAAACATTCGCTTATATCGCAACTTTC 2934
1208 LeAlaLysValAsnGlnSerAlaThrGlnGlnGlyAlaThrIle 1223
2935 GGCTACCCGACGACAAATTTGAACTGCGGAAAGTTCCGAAAGCACTTA 2984
1224 .....AlaSerAspLysVal 1229
2985 CACTTGGCGGTCAACAATACCGCAACGACCTGCAAGCTCGAACAAT 3034
1229 LThrLeuSerAlaAsnGlyLysAspAla..... 1240
3035 TGACGGTAGTGGAAGAAAGACAAACACCGCTGTCGAAACCTTAAT 3084
1240 ..... 1240
3085 TTCACCTGCAAAACGACACGTCGATGCGCGCGCTGCTTACCACT 3134
1241 LeuHisLeuGlnGlyAlaLysValSerGlyGlySer.....AlaAlaLe 1255
3135 CATCCGCAAGACGCGAGTTCGCTCGCATTAATCGGTCAAGAACAG 3184
1255 uGlnAlaLysAsnGlyGlyIleLeuLeuGlnSerAlaLysAsnGlnGln 1272
3185 AGCTTTCGAC...AACTCGC...AAGCAGAACCCAA.....AAA 3222
1272 IsLysAspAsnThrSerLeuGlyIleLysAlaAsnAlaLysGlyGln 1288
3223 CAGCGGAAAGAAC..... 3239
1289 ThrPheAsnLysAspAlaGlyLysValAspProAsnThrGlyLys 1305
3240 CGGCAAAAGCTTTCACGCGCTGATTCGCGCGCGGCGGATCGCGGAA 3289
1305 pRhnHisThrLeuGlyAlaGlyLeuLysValGlyAlaGlnGlnAsp 1322
3290 AACAGAAAGCGTTGCCGAAACGCGCGCGGAGGAGCGGAAATGTC 3339
1322 YThrThrHisAla.....AsnThr 1328
3340 GGCATTAAGCGCGGAGGAAAGAAAGGAGTGCAGCGGATTAAGA 3389
1329 GlyIleThrAlaGlyAsp.....ValThrLeuAsnSerGlyLysAs 1342
3390 CACGCGCTTGGCGAAACAGCGC...GAAGCGAAACCGCGCGCTCA 3436
1342 PThrArgLeuAlaGlyAlaLysValAspAlaAspSerValGlnGlyLys 1359
3437 CGGCTTCCCGCGCGCGCGCGCGCGCGGATTCGCGCACTGCA 3486
1359 aGlyGlyAspLeuHisValGlnSerArgLysAspAla..... 1371
3487 CCCCAACCGAGCCCAACCGCAGCGCATGATCAGCGGTTAGCCAA 3536
1372 .....GluAsnGlyValLysValAspValAs 1380
3537 TAGCGTTAGTAGAATTTCCGCCACGCTCAACAGCTTTCCCGCATC 3586
1380 PAlaGlyLeuSerHisSerAsnAspProGlySerSer..... 1392
3587 AGAGCAAAATTAAGCCCGCTATTTCGGAAGACCGCGCAAGCGCTTGG 3636
1393 .....IleThrSerLysLeuSerValGlyThrProArgTyrAl 1406
3637 ACAAGCGATCCGGGACCAACACTACGTTTCGCAAGATTTCCGCGC 3686
1406 aGlyLysValLysGlyLysLeuGlnAlaGlyValAsnLysValAlaAsp 1423
3687 CTACCGCAACAAACGACCTGCCCAATCGATGACGAA..... 3730
1423 LaThrThrAspLys.....TyrAsnSerValAlaArgArgLeuAsp 1436
```

```

3731 .....ACCTGGCAGCGCGCGCTGGCATCTGTTTCG 3765
FT REPEAT 101 106 1-1.
FT REPEAT 205 210 1-2.
FT REPEAT 279 284 1-3.
FT REPEAT 410 415 1-4.
FT REPEAT 635 640 1-5.
FT REPEAT 674 679 1-6.
FT DOMAIN 149 172 4 X 6 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 149 154 2-1.
FT REPEAT 155 160 2-2.
FT REPEAT 161 166 2-3.
FT REPEAT 167 172 2-4.
SO SEQUENCE 928 AA; 101228 MW; 67DA5466AB92390 CRC64;

1470 rpaaparglyAlaargThValglyValalValalysaspsertleThr 1486
3797 GCATCGGCAACTGGCAGCG...TTGCCAGCGCGCGCTTTCGGGCA 3843
1487 GlyProAlaGlyArgGlnGlnHisLeuValAsnAlaAspValValAs 1503
3844 TACGGCATCGACGAGTTCTACATCGCATCGCG...GGCGCG 3886
1503 nasnasAlaValAlGlyGlnGlnSerAlaLeAlaGlyLysAsnGlyVala 1520
3887 GTTTACGAGCGCGCGCTTCAGACGCGCATCGAGCAAAATCCGC 3933
1520 lalenglnValAlGlyGlnGlnThrGlnLeuThrGlyGlyGluLeuArg 1535
seq_name: SwissProt_40: HXA2_HAEIN

```

```

seq_documentation_block:
ID HXA2_HAEIN STANDARD; PRT; 928 AA.
AC P4354;

```

```

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme/hemopexin-binding protein precursor (Heme:hemopexin utilization
DE protein A).
GN HXA2.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
OC Haemophilus.
RX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-35.
RC STRAIN-DL42 / SEROTYPE B;
RA MEDLINE=95115556; PubMed=7815944;
RA Cope L.D., Thomas S.E., Latimer J.L., Slaughter C.A.,
RA Mueller-Eberhard U., Hansen E.J.;
RT "The 100 kDa heme:haemopexin-binding protein of Haemophilus
RT influenzae: structure and localization."
RL Mol. Microbiol. 13:863-873(1994).
RN [2]
RP SEQUENCE OF 1-30 FROM N.A.
RC STRAIN-DL42 / SEROTYPE B;
RA MEDLINE=95270579; PubMed=7751272;
RA Cope L.D., Yogev R., Mueller-Eberhard U., Hansen E.J.;
RT "A gene cluster involved in the utilization of both free heme and
RT heme:hemopexin by Haemophilus influenzae type b."
RL J. Bacteriol. 177:2644-2653(1995).
CC -1- FUNCTION: BINDS HEME/HEMOPEXIN COMPLEXES.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U08348; AAA74138.1;
CC Transport; Signal; Repeat.
FT SIGNAL 1 21
FT CHAIN 22 928 HEME/HEMOPEXIN-BINDING PROTEIN.
FT DOMAIN 101 679 6 X 6 AA APPROXIMATE REPEATS.

```

```

FT REPEAT 101 106 1-1.
FT REPEAT 205 210 1-2.
FT REPEAT 279 284 1-3.
FT REPEAT 410 415 1-4.
FT REPEAT 635 640 1-5.
FT REPEAT 674 679 1-6.
FT DOMAIN 149 172 4 X 6 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 149 154 2-1.
FT REPEAT 155 160 2-2.
FT REPEAT 161 166 2-3.
FT REPEAT 167 172 2-4.
SO SEQUENCE 928 AA; 101228 MW; 67DA5466AB92390 CRC64;

alignment_scores:
Quality: 208.50 Length: 1107
Ratio: 0.419 Gaps: 59
Percent Similarity: 44.986 Percent Identity: 19.874

```

```

alignment_block:
US-09-303-518D-649 x HXA2_HAEIN

```

```

Align seg 1/1 to: HXA2_HAEIN from: 1 to: 928

```

```

454 GGCATCCCTTATGGCGCGCATATCATATGCGGTTGCATAATTTGT 503
17 GlyAlaAlaValAlaSerThrProAspPheGlnHisHisIsthrVa 33
504 CACAGATGCGACGACCTGTGAATGACCACTATATGATGGCGCAAT 553
33 IphGlyThrValThrIleGlyIsthr..... 43
554 ATATGATCAAAATATATACCTGACGCTGTCTGTTGGCGAGCGAG 603
44 .....AlaAspLysMetThrIleLysGlnGlySer 53
604 CATAT.....TGGCATCTGATGAA..... 624
54 AspLysAlaGlnIleAspIrpLysSerPheAspIleGlyGlnLysG 70
625 .....GATGAGCCCATATACCGCGAAAGTTCATATCATATTCGAA 664
70 uValLysPheGlnGlnProAsnGlnHis.....Alav 81
665 GTCCGATTTCTGGCGCTGTGGCAATACCTTTCACAAATGCA... 711
81 alaIatyrAsnArgValIleGlyGlyAsnAlaSerGlnIleGlnLys 97
712 ...TCAGTGTGGCACAGCTCACTTATGCTAGTGAATAATTAACATAG 758
98 LeuThrAlaAsnGlyLysValTyrLeuAla.....As 108
759 CCCATATGTTTTTATCCACAGAGAGCTCA..... 789
108 nProAsnGlyValIleIleThrGlnGlyAlaGluLeuAsnValAlaG 125
790 .....TTGGCGAGCTGCTCA 807
125 euleAlaThrThrLysAspLeuGlnuArgIleSerGluAsnSerAnser 141
808 CCATGTTATCTATGATGCCCAAAAGCAAAAGTGAATTAATGAG... 855
142 TyrGlnPheThrArgValThrLysAspArgGlnValLeuLysGlnG 158
856 GTATATGCAACGGCG.....AACCCCTATATGGAAGAA 889
158 uValLeuLysAspGlyGlnValAlLysGlnGlnValIleAsnGln 175
890 GCAATGCTTCCAGCTGTGCTGAAGATTGTTCTAT.....GAT 930
175 Lysn.....IleThrAlaGlnAspPheValValLeuAsnGlyAsp 188
931 GAAATCTTGTGCGAATACCATTCAGTATTTACAGAACACAGCTCAA 980

```


189 G L V A I L L E A S N L Y S G I A S N I L E A S N V A L G I L Y S A S E R T H R I L E A S |||
189
981 T G G A A T A C T C T T T A A G A C A T A T A T A T G C A C A 1017
205 N G I L Y S V A L T Y L E U S E R S E R L Y T Y R A S N P H E T H R P H E T H R L E U P R O A |||
1018 . . G G A A A A T C A A T G C C A A A C A T G A A C A C A T T C T C G C T A A T G A T T A |||
222 S P E R I L Y L E S E R V A L A L E U G I A S P A S N T H R V A L G I N C L Y L E A L |||
1066 A A A C A C A G A . . A C G T T C A T T G T T A T G T T T T T T A T C C G A C A C G |||
239 L Y S A S N G I U C L Y S E R I L E Y S A L A G I Y G I U L I E T H R L E U S E R A L A L Y S G I |||
1113 A A G A C A C C T G T T T A T C A T G T C A G G T G T G C A C A |||
255 Y A R Y S G I N A L A L E U A S P S E R L E U V A L M E T A S N A S N C L Y V A L L E U G I A |||
1151 G T T A T C G A C C C A C T G A A T A T G A G A A A A T |||
272 L A T H R L Y S V A L S E R A S N L Y S A S N G I L Y S V A L V A L L E U S E R A L A A S P A S N |||
1183 A T T T C C T T T A T T G A C G A A G A A A A G C G A T T G A T A 1218
289 V A I G I U L E U A S N A S N G I L S E R A S N I L E L Y S G I U L I E V A L T H R P H E G I |||
1219 C T T A C C A G C A A C A T C A A T C A A G T G C T G A G A T T A T |||
305 Y A L A S P A V A L T H R S E R A S N L Y S G I U L E U Y S A S P A S N I L E L Y S I L E T H S |||
1256 A T T T C A A G A G A T T T A C G T C T C G C C T G A A A T A A C G A A A T T G G C A A |||
322 E R L Y S T H G I Y S E R L Y S E R V A L T H R S E R P R O L Y S I L E A S N P H E T H R 336
1306 G G C G G G G C G T C A T A T C A C T G A A G A C A G T A C C T |||
337 G L Y S E R V A L A S N I L E A S N G I L Y A S N P H E G I Y A R G I U A S P E R T H R T H |||
1341 T A C T T G G A A A G T A A C G C G T G C A A A C G A C C C C T G C C A A A T C G C G . 1389
353 T H S T Y L Y S A S P E L U P H E L Y S L Y S L E U A S N T |||
1390 A A A G C A C G C T G C A C G T T C A A C C |||
364 H R G I U V A L A S N I L E A S P V A L P R O A S P A S N G I U A S N I L E A R G I L E A L A S P |||
1414 A A A G G G A A A C C A G C T C A T C A G C G G C G A C G A C G T A C A T T T T |||
381 I L E G I A S P A S N T H R G I Y T H R G I Y T H R G I Y T H R 394
1464 G G A T C G A C G A C A C G A T A A A G C A A A A C A C C T T T A G T G A A A T C G |||
395 S E R S E R P H E L I E G I N T H R G |||
1514 G C T T G T C A C G C G A G G G T A C G G T C A C A C T A A T G C C |||
401 L Y A L E U S E R S E R L E U E A L A S N A S N G I L Y S V A L A S N L E U L Y S G I |||
1552 G A T A T C A C T T C A C C C G A C A A C T A T T C . . . G G T T T C G C G C G . . 1596
418 A S N A S N V A L A S N I L E S E R G I Y A R G I L I E H I S I L E A S P S E R P H E A T G L Y S E |||
1597 G A C G T T T G G A T T T A A C G G C |||
434 T A S P S E R L E U L Y S I E U T H R A S N L Y S G I Y H I S I L E A S P I L E A S N A |||
1619 A T T G C T T T C T T C C A C C G T A T T C A A A T A C C G A T G A A G G G G A T G A T T |||
451 L A S P I L E H S S E R L Y S G I Y A R G L E U P E |||
1669 G T C A C C A C A A T C A A G A A A T C C A C G T T A C C A T |||
1706
461 P H E I L T H R S E R L E U G I A S N G I U G I U A S P H E L Y S S E R A S N I L E T H R I L |||
1707 T A C A G C A A T A A A G A T A T T G C T A C A C C G C A A T A T A C A C A |||
477 E T H R A S P S E R L Y S I L E A S N L E U G I Y A S N G I L A M E T G I L E U G I Y A R G S |||
1748 G C T T G A T G A C A A A A A A A A T T G C C T A C . . . A A C G T T G C T T T G C G A G |||
494 E R V A L A S P I L Y S A S P T Y R A S P A S N A R G T R P 504
1795 A A A G A T A C G A C A A A A C G A C C G G C G |||
505 G I N L Y S T H R G I U C L Y S E R I N A R G L Y S L Y S P H E A P V A L Y |||
1821 G C T C A A C C T T G T T T T C A C A C C C C G C G A A A C C C G C T G C T T T |||
518 S M E T E R S N V A L G I U P H E A S N G I N V A L A S P A S P V A L I L E U A |||
1871 C C G C G G A A C A A T T T A A C G C A C A T C A C G C A A C A A C |||
533 I A G I Y G I P H E G I U L Y S V A L A S N L E U A S P 542
1912 G G C A A C T G T T T T C A C G C G C A C A C C G C A C G C T A C A T G A T T T |||
543 L Y S I L E V A L A T H R G I Y G I N 549
1962 A A A C G A C C A T T G T C G C A A A A A G A G G C A T T C C T C G G G G A A A T G C T G T |||
549 549
2012 G G G A C A A C G A C T G A T A C C C A C A T T T A A A C G G A A A C T T C C A A T T |||
550 T H R S N P H E T H R I L E |||
2062 A A A G C G G A C A G C G G T G T T C C C G A A T G T T G C C A A A G T G A A A . . G G |||
555 A S P E L Y G I L Y V A L S E R A R G A S N G I Y A R G L Y S T Y R G I U L Y T G I |||
2109 G A A T T G C A T T T G A C A A T C A C C C C A C A 2139
568 Y V A L L E U A S P L E U A S P L Y A R G T H R C I L E U S E R G I U L E A N G I N G I L Y A |||
2140 G T T T T G G T G C A C C C C A C C A T C A A A C C A C A C A |||
585 R G A R G A R T R P G I Y T Y R T Y R A S P L E U G I U L E U A S P M E T A S N A R G A L A |||
2173 A T C T G A C A C G T T T G A C T G A C G G G T C T G A C A A A T T G T G C A A A A A C |||
602 T Y R L E U T Y A R G P H E A S P L E U P H E A L A T H R L Y S A S N T H R G I Y A R G S E R T H |||
2223 C A T T A C C G A C 2232
618 I L L E L Y S A S P T H R G I U L I L E A S N I L E S E R A S N S E R A S N I L E A S N L E U L Y S A |||
2233 G A T A A G A T T G C T T C A T T G A C P A A G |||
635 S N G I L P H E V A L A I S L E U L E A L A G I U L Y S I L E L Y S L E U A S P A S N S E R L Y S |||
2260 A C C G A C A T C A C G G C A A T G T C G A T C T T G C C A T C A |||
652 I L E A S P I L E T H R P H E A S P L Y S A S P A S N S E R G I N A S P H R L E U A L A G I N T H |||
2295 C G T C A T T A A T C T C A C A G G C T T G C A C A C T C A A C G C A A T G T T A |||
668 T A S N A R G L E U G I Y M E T A S N G I L Y S V A L S E R M E T I L E A S N H I S I L E I L E |||
2342 G T C C A A A T G C G A T A C A C G T T A T A C A G T C A C C A C A C C A C C C A A A C |||
685 Y S I L E V A L G I A S P G I U L Y S G I U G I Y L I E S E R P R O T H R G I Y T H R T Y R A L A |||
2392 G G C A C C T T A C C T C G T G G C A A T G C C C A C A C A A C A T T T A |||
702 T H R M E T P H E U L I E G I U L E U I L E G I U L Y S S E R S E R I L E P H E V A |||

```

2433 TCAAGC.....ACATTAAAGGCAACATCGCCTT 2464
      :::::::::::::::
718 IlySerHisGlnGlyThrPheLysThrAspGlyAsnThrLysIleA 735
      :::::::::::::::
2465 CGGCAATGCTTCA...TTTAATCTAAGGACACCGCGTACAAACGCG 2511
      :::::::::::::::
735 IagLysTyrSerLysGlnAspLysIleThrAlaIleAsnThrGly 751
      :::::::::::::::
2512 AGTCGACGCTTTCGGCAACGCTAAGCAACGTAAGCCATTCCGCACT 2561
      :::::::::::::::
752 .....GlyArgAlaIleGlnIleVal.....LeuI 760
      :::::::::::::::
2562 CAACGTAATGCTTCCTCCTAGCGGATATGCGATTTCCATTGGAAGCA 2611
      :::::::::::::::
760 eAsnGlyAlaLeuGlySerAlaAspAsnAspAla...AsnIleAlaAsn 776
      :::::::::::::::
2612 GCGCGTTTACGACAAATCAGCGCGGCAAGATACGCGATTAACCTTA 2661
      :::::::::::::::
776 eAlaPheThrIleGlyAspSerAlaAsnThrLysThrIleGlnAsn 792
      :::::::::::::::
2662 AAAGACGGAATGAGCCTGCGTCAAGCAGCAAGATTAAGCATTTAA 2711
      :::::::::::::::
793 AlaAspIleThrAlaLeuAlaProAsnGly.....GlyThrAlaTyr 806
      :::::::::::::::
2712 CCTTGCAACGCGCACCATTACATC.....AATTCGCGCTATC 2749
      :::::::::::::::
806 rLeuSerSerLysAspValGluIleGlnValLysProAsnSerAsnPe 822
      :::::::::::::::
2750 GCCACGATCGCGCAGGCGCAACCGGACGACAGATGCGCGCGCG 2799
      :::::::::::::::
823 .....ThrPheGlnLeuProArg..... 829
      :::::::::::::::
2800 CGCGGTTCCGCGCGTCCGCGCTTCCTATTATCCGTTACACGCGCA 2849
      :::::::::::::::
829 ..... 829
      :::::::::::::::
2850 TTGCGTAGAATCCGTTTCAACAGCTGACGGTAAACGCG.....A 2890
      :::::::::::::::
830 .....GluLysAsnLeuAsnGlnThrLysIleAsnGlyAlaSerThr 844
      :::::::::::::::
2891 AATTGACGCTCAGGCAACATTCGCTTATGTCGAGACTCTTCGCGTAC 2940
      :::::::::::::::
844 yLysLeuSerGluArgGlyPheAlaArgLeuTyrAspLysIleAsnGly 860
      :::::::::::::::
2941 CGCAGCACAATTAAGCTGGCGGAAGTTCGGAAGCATTAACACTT 2990
      :::::::::::::::
861 ArgAlaSerAsnLeu..... 865
      :::::::::::::::
2991 GCGCGTCAACAATACCGCAACGCTGCAAGCTCGAACAATATGACGG 3040
      :::::::::::::::
866 .....SerAlaGlnGlnLeuAsnV 872
      :::::::::::::::
3041 TAGTGAAGAGAAAAGACAAACACCGCTGCCGAAACCTTAATTCAC 3090
      :::::::::::::::
872 alThrAspAlaSerGln...LysIleIleAsnThrLysLeuValSerSer 887
      :::::::::::::::
3091 CTGCAACGCAACACGCTC.....GATCGCGCGCGCTG 3122
      :::::::::::::::
888 LeuAspValGlnLysLeuValSerValAlaValLysAspAlaGly 902
      :::::::::::::::
3123 GCGTTACCAACTCATCCGCAAGACGGCGAGTCCGCTGATATATCGG 3172
      :::::::::::::::
903 .....AsnGlyC 905
      :::::::::::::::
3173 TCAAGACAAAGAGCTTCCGCAAA.....CTC 3201
      :::::::::::::::
905 ySeGlnIleGlnGlnPheGlyAspLysGlyAsnAsnThrLysValSerVal 921
      :::::::::::::::
3202 GGCAGGCAAGACCAAAAA 3222
      :::::::::::::::
922 GlyGlnLeuGlnAlaGlnGln 928

```

```

seq_name: SwissProt_40:WAPA_BACSU
seq_documentation_block:
ID WAPA_BACSU STANDARD; PRT; 2334 AA.
AC 007833;
DT 01-OCT-1994 (Rel. 30, Created)
DR 01-OCT-1994 (Rel. 30, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Wall-associated protein precursor.
GN WAPA OR N176.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93302506; PubMed=8316082;
RA Foster S.J.;
RT "Molecular analysis of three major wall-associated proteins of a
RT Bacillus subtilis 168: evidence for processing of the product of a
RT gene encoding a 258 kDa precursor two-domain ligand-binding
RT protein."
RL MOL. Microbiol. 8:299-310(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=95219088; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapa loci."
RL Microbiology 141:337-343(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=97124196; PubMed=8969509;
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA Miwa Y., Fujita Y.;
RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT containing the hlc and cel loci, and creation of a 177 kb contig
RT covering the gut-saxx region."
RL Microbiology 142:3113-3123(1996).
CC -1- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
CC MOTILITY, SECRETION OR DIFFERENTIATION.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
CC INTO THE MEDIUM.
CC -1- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
CC MOTIF REPEATED 31 TIMES.
CC -1- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
CC SIMILARITY TO THE REPEAT IN E.COLI RHB GROUP OF PROTEINS (RHB-A-D).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation on
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L05634; AAA22883.1; -
DR EMBL; D31856; BAA06656.1; -
DR EMBL; D29985; BAA06260.1; -
DR EMBL; D83026; BAA11683.1; -
DR EMBL; Z99124; CAB15959.1; -
DR PIR; S32920; S32920.
DR Subtilisin; BG10797; WAPA.
DR InterPro; IPR003305; CBD_6.
DR Pfam; PF02018; CBD_6; 1.
DR Cell wall; Repeat; Signal; Complete proteome.
KW SIGNAL 1 28 OR 32 (POTENTIAL).
FT CHAIN 29 2334 WALL-ASSOCIATED PROTEIN.

```

```

FT DOMAIN 504 869 3 X 101 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 504 605 1-1.
FT REPEAT 636 736 1-2.
FT REPEAT 769 869 1-3.
FT DOMAIN 1021 2139
FT REPEAT 1021 1040
FT REPEAT 1042 1061 2-1.
FT REPEAT 1063 1082 2-2.
FT REPEAT 1083 1102 2-3.
FT REPEAT 1109 1128 2-4.
FT REPEAT 1129 1148 2-5.
FT REPEAT 1150 1169 2-6.
FT REPEAT 1174 1193 2-7.
FT REPEAT 1199 1218 2-8.
FT REPEAT 1219 1238 2-9.
FT REPEAT 1267 1286 2-10.
FT REPEAT 1466 1485 2-11.
FT REPEAT 1690 1709 2-12.
FT REPEAT 1711 1730 2-13.
FT REPEAT 1732 1751 2-14.
FT REPEAT 1753 1772 2-15.
FT REPEAT 1795 1814 2-16.
FT REPEAT 1820 1839 2-17.
FT REPEAT 1840 1859 2-18.
FT REPEAT 1861 1880 2-19.
FT REPEAT 1887 1906 2-20.
FT REPEAT 1908 1927 2-21.
FT REPEAT 1929 1948 2-22.
FT REPEAT 1969 1988 2-23.
FT REPEAT 1983 2002 2-24 (APPROXIMATE).
FT REPEAT 2008 2027 2-25.
FT REPEAT 2028 2047 2-26.
FT REPEAT 2051 2070 2-27.
FT REPEAT 2071 2090 2-28.
FT REPEAT 2093 2112 2-29.
FT REPEAT 2120 2139 2-30.
FT REPEAT 2139 2158 2-31.
SO SEQUENCE 2334 AA; 258329 MW; B75138CCD27BBA3 CRC64.

```

alignment_scores:
 Quality: 208.00 Length: 1236
 Ratio: 0.382 Gaps: 63
 Percent Similarity: 44.013 Percent Identity: 19.822

alignment_block:

US-09-303-518d-649 x WAPA_BACSU

Align seg 1/1 to: WAPA_BACSU from: 1 to: 2334

```

160 TATCCGACTTTCGCGAAATTAAGCAAGTTTCAGTGGCGGGAAGA 209
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
988 TTTLeuGluLeuThrGluThrAlaSerPheLeuLeuLeuThrLys 1004
210 T.....
1004 polrhrsmalatyrrheasnlysllysGlyGlyLysLeuGlnLysVal 1021
211 .....ATTGAGTTTACACAAAGGAGGAG 237
1021 aLasrGlyHisAsnAsnAlaThrValTyrrhrTyrrAsnAspLysVal 1037
238 TTGGTCGCGAAATGACATGACAAAGCCCGATGATTTCTGTGTGT 287
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
1038 LeuThrAlaLeuThrAspAlaSerGlyArgLysLeuThrPheThrLys 1054
288 GTCCGCGTAAAGCGCGTGGCGCATTTGGTGGC..... 318
    : : : : : : : : : : : : : : : : : : : : : : : :
1054 pGluAsnGlyHisValThrSerLeuThrGlyPheLysAsnLysVal 1071
319 .....GATCAATATATTGAGCGTGGCAGAT...AACGCG 351
1071 hrTyrrSerTyrrGluAsnAspLeuLeuLysValThrAspThrAsp 1087

```

```

352 GCGTATTAACACGTTGATTTGGTCGGAAGAAATCCCGATCAACA 401
    : : : : : : : : : : : : : : : : : : : : : : : :
1088 ThrValThrSerTyrrAspLysAspSerGluGlyThrLeuValLysGln 1104
402 TCGT.....TTTACT..... 411
1104 rSerLAsnSerThrGlnAlaLysProValPheThrGluThrGlnLys 1121
412 .....TATTAATTTGTGAACGAAATATTATAAGCAGCGACTTAAAGC 456
1121 ergLynHisArgLeuGlnLysAlaLeuAsnAlaLysGluThrVal 1137
457 CATCCTTATGGCGCGAT.....TATCATATGGCGGCTTT 491
1138 TyrrSerTyrrAspAlaSerLysLysThrLeuLeuMetThrGlnProAsn 1154
492 GCATTA.....TTTG 502
1154 YArgLysValGlnTyrrGlyThrAsnGlnAlaGlyAsnProLeuGlnVal 1171
503 TCACAGATGCGAAGCCGTGGAATGACCGATTATGAGTGGCGGAGA 552
    : : : : : : : : : : : : : : : : : : : : : : : :
1171 LeAspAspAlaGluGlyLeuLysIleThrThr.....AsnThrLys 1184
553 TATATGATCAAAATTAATTAC.....CTGACCGGTGTGC 587
1185 Tyr...GluGlyAsnAsnValValGluAspValAspProAsnAspVal 1200
588 TATGGGCGGAGCGAATATGGCGATCTGATGAGAGTGGCCCAATA 637
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
1200 YThrGlyLysAlaThrGluSerTyrrGlnTyrrAspLysAsp.....GlyA 1215
638 ACCGCGAAAGTTGATATGATATGCAAGTGCCTATCTTGGCTGTGT 687
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
1215 snValThrSer.....ValLysAspAlaTyr.....Gly 1224
688 GGCATATACCTTGCACAAATATGATCAGTGGTGGCAGATCAACTAG 737
    : : : : : : : : : : : : : : : : : : : : : : : :
1225 ThrGluThrTyrrGluThrAsnLysAsnAsnAspValThrLysMetLys 1241
738 TAGTGA.....AAATTAACATAGCCCATATGTTTATACACAGAG 784
    : : : : : : : : : : : : : : : : : : : : : : : :
1241 PThrGluGlyAsnValThrAspIleAlaTyrrAspGlyLeuAspAlaVal 1258
785 GGTCAATTTGGCGACAGTGC.....TCACCAATGTTATCTATGAT... 825
    : : : : : : : : : : : : : : : : : : : : : : : :
1258 ergLutThrAspLeuSerGlyLysSerSerSerAlaValTyrrAspLys 1274
825 .....
1275 TyrrGlyAsnGlnIleGlnSerSerLysAspLeuSerAlaSerThrAsn 1291
826 .....GCCCAAAAGCAAAAGTGGTATTAAT... 852
1291 eLeuLysAspLysSerPheGlnAlaGlnLysSerGlyTyrrAsnLeuTh 1308
853 .....GGGTATTGCAACGCGCAACCCCTAT 879
1308 lAserLysAspArgArgLysIleSerValIleAlaAspLysSerGlyVal 1324
880 ATAGAGAAAGCAATGCGTTCCAGCTGGTCTGAAGATGGTTC..... 924
    : : : : : : : : : : : : : : : : : : : : : : : :
1325 LeuSerGlySerLysAlaLeuGluValLeuSerGlnSerThrSerAla 1341
925 .TATGATGAATCTTTGCGAGATACCCATTGATTTCTAGAACAC 973
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
1341 yThrAspHisGlyTyrrSerSerAlaThrGlnThrValGluLeuGluPro 1357
974 GTCAAAATGGGAATACCTTTTAACGACGATAAATGCGACAGAGAAA 1023
    ||| : : : : : : : : : : : : : : : : : : : : : : : :
1358 .....AsnThrThrTyrrThrLeu.....SerGlyLys 1366

```

1024 ATCAATCCCAACATGAAACAATCTCTGCTTAATAGATTAAACACG 1073
|||||
1367 lIeysThr.....AspleuAlalysSerAr 1375
1074 AACCGTCATTTGTTAATGTTCTTATCCGACAGACAGCAAGAACCTG 1123
|||||
1375 gAla.....TyrPheAsnIleAspleuArGAspLyAspGlnIlyArG 1390
1124 TT.....TATCATGCTCCAGCTGCTGCATCAACATTTAT 1155
1390 legIInPrIleIleAsnGlnIlySerAlaLeuAlaGlnIlyAsnAspTrp 1406
1156 CGACCCAGA.....CTGAATATGAGAAATAT 1184
1407 ThrIyAsnArGlnIleThrPheThrProAlaAsnAlaGlnIlyAsAla 1423
1185 TTCCCTTATGAC.....GAAGCAAAAGCGAATGCA 1216
1423 lValIyMetGlnValAspHisIlyAspLysAspGlnIlySalAT 1440
1217 TACTTACCAGC...AACATCAATCAAGGTGCTGAGGATATATTCGAA 1263
1440 rPheAspGlnValGlnLeuGlnIlySgIly.....Glu 1450
1264 GGAGATTATACGCTCGCTGAAATTAACGAACTTGCGACGCGCG 1313
1451 ValSerSerSerTyrAsnProValGlnAsnSerPheThrSerAla.. 1466
1314 CGTTCAATACATGAAAGACAGTACGCTTACTGAAATGAACGCGG 1363
1467ThGlnsInTrpAsnValSerGlnAlas 1476
1364 CA.....AACGACCGCTGTCGCAAAATCGGCAAA 1392
1476 eValAspSerGlnIlyGlnPheAsnAspAspValSerLeuIlyAla 1492
1393 GGCACGCTGCACGCTCAAGCCAAAGGCGAAACCAAGGCTGATAGCGT 1442
1493 ArgThrSerAlaSerGlnAlaGly.....SerVa 1502
1443 GGCAGCAGGTGATCATTTTGGATGAGCAGCAGACGATTAAGCAAAA 1492
1502 lThrIySgInThrValIlyValLeuGlnIlySerAlaAsnAspLys..... 1516
1493 AACACGCTTATGAAATCGCTTGTCACGCGGACGCGTACGTCGAA 1542
1516 1516
1543 CTGAATGCCGATATACATTCAACCCGCAACCTATTTGGCTTCG 1592
1517ProValIlyr..... 1519
1593 CGCGGACGTTTGATTTAAAGCGGCTTGGCTTCCACCGATATTC 1642
1520LeuThrLeuThrGlyMetSer..... 1526
1643 AAAATACGATGAAGGCGCATGTGTCAACACATCAAGCAAGAA 1692
1527LysAla 1528
1693 TCCACCGTTACCATTAAGCAATTAAGATTTGCTACACCGGCAATTA 1742
1529 SerSerValIlySpherThrAspGlnIlyAspIlySerLeuGlnAlaAsnVa 1545
1743 CAACGAGCTTGATGACAAAGAAATGCTTACACGCTTGCTTT... 1788
1545 lThrIyAlaAspGlnIlySerThrGlyIle...TyrAsnAlaIlySpher 1561
1789 ..GGCGAGAAAGAT.....ACACGCAAA 1809
1561 eGlnIlyThrGlnIlyTrpAsnArGAlaAlaValAlleProIlyThrIys 1577
1810 ACGAACGGCGG.....CTCAACTTGTATTACAGCCGCGGAGAGAA 1853

1578 ProIleAsnIlyValAspIleSerIleLeuPheGlnIlySerAlaThr.. 1593
1854 CCGCACCTCTGCTGCTTCCGCGGAAACAATTTAAACGCAACATCAGC 1903
1594GlyThrValIlyrPheAspIleAlaGly 1603
1904 AACCAACGGCAACCTGTTTTCAGCGGCAACACACCGGACCTTAC 1953
1603 euIleGlnIlySerLeu.....ThrIySerThrIy 1614
1954 AATCATTTAAACGACCACTTGTCGCAAAAGAGGCAATTCCTCGGGGA 2003
1615 AspSer...AsnGlnAsnIlyValIlyThrIySgInIlyAspGlnIly 1629
2004 AATGCTGTGGACACAGCAGTGCATGATCAACCGCATTTAAAGCGAAACT 2053
1630 .TyrAlaThrSerThrAspIlyTrpGlnIlyIlyIlyIlySerGln 1646
2054 TCCAAATTAAGCGGACAGCGGCTGCTTCCGCAATGTCGCAAAAGT 2103
1646 hrAspAlaIlyGlnIlyIlyThrIyThrIyThrIyAspGlnAlaAspGln 1662
2104 AAAGCGATTGCGATTGACGATCAGCCCAAGCAATTTGCTGTCG 2153
1663 LeuThrAsnMetThrLeuSerAsn..... 1670
2154 ACCGATCAAGCCCAACATCTGTACAGCTGCGACTGACGGGTGCA 2203
1671GlyThrSerIleLeuHisSerTyrAspGlnIly 1682
2204 CAATGCTGTGCAAAACATTAAC... 2229
1683 ..AsnGlnIlyAspValSerIlyThrIleArGAlaGlnIlyAspGlnIlyIly 1698
2230 ...GACGATTAAGTATGCTTCAATGACTAG...ACGACATACGCG 2273
1699 rPheGlnIlyAspValMetGlnIlySerLeuValIlyThrIlyAspProIle 1715
2274 CAATGTGATCTTCCGATCAC...GTCATTAATTAATGCAAGGCTTG 2320
1715 yAsnValIleuAlaSerGlnIlyAspAlaAsnSerIleuThrIlyIly 1732
2321 CCGACACTCAAGCGCAAT.....CTTACTGCAATGCGATACAGCT 2361
1732 leSerProAsnGlnIlyAsnGlnIlyValSerLeuSerTyrAspIlyThrAspArG 1748
2362 TAT...ACAGTCAAGCCACAGCCACCAA..... 2388
1749 ValIlySerIlySerIlyAsnGlnIlyThrIlyIlyIlyIlyIlyPheThrIlyAs 1765
2389 ...AACGGACCTTACGCTGCGGCAATGCCAAGCAACATTTAATC 2434
1765 pLysAsnGlnIlyAsnGlnIlySerValIlyAsnIlySgIn.....G 1778
2435 AAGCCACATTAACAGGCAACATCGGCTTGGGCAATGCTTAT 2484
1778 lAsnIlyThrIlyIlyAsnGlnIlyThrPheAspAsnIlyAsnGlnIlyIly 1794
2485 CTAGCGACCGCGGCTTACAAAGCGCATCTGACG..... 2520
1795 LeuThrAsp.....ArgGlnIlySerGlnIlyThrIlyrProse 1808
2521CTTCCGCGACAGCTTAAG 2539
1808 rAspSerAspLysLeuIlyThrPheSerTrpIleHisGlnIlyAspGlnIlySg 1825
2540 CA.....AACGTAAAGCATTCGCGACTCAACGATTAATGTCCTAGCC 2583
1825 lYThrAsnGlnIlyPheThrIlyAsnIlyLeuAspGlnMetIleGlnMetIly 1841
2584 GATAAGCA.....GTAATCCATTTGAAGACCGGCTTTACCGG 2624

```

1842 AsperTherSertYrSerPheAspTyrAspGluAsnGlyAsnValGI 1858
2625 ACAATATCAGCGCGCAGAGATACGGCATTTACACTTAAAGACAGCGAAT 2674
1858 nThrPheIleThrGlyAsnGlyValGlyThrSerPheSertYrAspGluA 1875
2675 GGAGCGTCCGCGTACGAGCAGGAAATTAAGCAATTTAACTTGACAAACCC 2724
1875 rGAsnLeuValSerSertLeuHisIleGlyAsp.....LysAsnGly 1888
2725 ACCATTATACACTCAATTCGGCGCTATGCGCAGCATGCGGAGGCGCAAC 2774
1889 GlyAspIleLeuThrGlySerGlyValGlyThrAspAlaAsnGly..... 1902
2775 CGGCACTGCGACAGATGCGCGCGCGCGCTTGGCGCGCTTGGCGCGCT 2824
1903 ..... 1905
2825 CCTATTATCGTTCACCGCCCACTTGGTGAATCCGCTTTCACACG 2874
1905 hrThrIleAsnSerSertAlaSerGlyValGlyThrGlyTyr..... 1919
2875 CTGACGGTAAACGGCAATTGAC.....GCTCA 2903
1920 .....GlyLysLeuAsnGlnLeuValLysGlyThrHisGluAs 1932
2904 GGAGACA...TTCCGCTTTATGTCGGAACCTTTCGGCTACCGGAGCGACA 2950
1932 polYThrValIleGlyThrGlyThrAspGlyPheGlyAsnArgLysThrV 1949
2951 AATTGACCTGGCGGAAGTTCCGAGGCACTTACACCTTGGCGGTCAAC 3000
1949 alThrThrIleLysAspGlySerSertLysThrValAsnAlaSerPheAsn 1965
3001 AATTACCGGACAGCACTGCAACCTTGAACAATTGACGGTGAAGAAG 3050
1966 .....IleMetAsnGlnLeuThrLysValAsn.. 1974
3051 AAAAGACACAAACCGCTGCGCAAAACCTTAATTTCACCTTGCAGAAACG 3100
1975 .....AspGluSertIleSertYrAspLysAsnGlyA 1985
3101 AACACGTCGATCGCGCGCGCTTACCAACTCATCGCAAGAGCGG 3150
1985 smrArgThrSerAspGlyLysPheThrTyrThrTPAspAlaGluAspAsn 2001
3151 GAGTTCGCGCTGATATCCGGTCAAGAAGACAGCTTTCGCAACAAT 3200
2002 .LeuThrAlaValThrLysGlyGluAspLysProPheAlaThrTyrL 2018
3201 CGGCAAGCGACAGCAAGCAAGAGCGCGCAAGAGACAGCGCAAGCGC 3250
2018 ySTyrAspGluLys...GlyAsnArgGlyIleGlnLysThrValAsnGlyLys 2033
3251 TTGACG 3256
2034 ValThr 2035

```

seq_name: SwissProt_40.PRTS_SERMA

seq_documentation_block:

ID PRTS_SERMA STANDARD; PRT: 1045 AA.

AC P09489;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Extracellular serine protease precursor (EC 3.4.21.-).

OS *Serratia marcescens*.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Serratia.

OX NCBI_TaxID=615;

RN [1]

SEQUENCE FROM N.A., AND SEQUENCE OF 28-35 AND 407-408.

```

RC STRAIN-IFO 3046;
RX MEDLINE=06223815; PubMed=3011754;
RA Yanagida N., Dozumi T., Beppu T.;
RT "Specific excretion of Serratia marcescens protease through the outer
RL membrane of Escherichia coli.";
RN J. Bacteriol. 166:937-944(1986).
RP [2]
RP PARTIAL SEQUENCE, AND PROCESSING.
RX MEDLINE=92348352; PubMed=1639760;
RA Shikata S., Shimada K., Kataoka H., Hornouchi S., Beppu T.;
RT "Detection of large COOH-terminal domains processed from the
RT precursor of Serratia marcescens serine protease in the outer
RL membrane of Escherichia coli.";
RJ J. Biochem. 111:627-632(1992).
CC -1 SUBCELLULAR LOCATION: Secreted.
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL, M13469; AAA26572.1; -
CC PIR, A29840; A29840.
CC HSSP: 099405; 1MPT.
CC
CC MEROPS: S08.094; -
CC InterPro: IPR000209; Peptidase_S8.
CC Pfam: PF00082; Peptidase_S8.
CC PRINTS: PR00723; SUBTILISIN.
CC PROSITE: PS00136; SUBTILASE-ASP; FALSE_NEG.
CC PROSITE: PS00137; SUBTILASE-HIS; 1.
CC PROSITE: PS00138; SUBTILASE-SER; 1.
CC KW Hydroxylase; Serine protease; zymogen; signal.
CC FT SIGNAL 1 27
CC FT CHAIN 28 645 EXTRACELLULAR SERINE PROTEASE.
CC FT PROPEP 646 1045
CC FT ACT_SITE 76 76 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 341 341 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC SO SEQUENCE 1045 AA; 112345 MW; 4924E50E4FF179C CRC64;

alignment_scores:
    quality: 207.50      Length: 1237
    Ratio: 0.382        Gaps: 65
    Percent Similarity: 43.897    Percent Identity: 19.563

alignment_block:
US-09-303-518d-649 x PRTS_SERMA ..

Align seg 1/1 to: PRTS_SERMA from: 1 to: 1045

1162 AGACTGATATATGCA.....GAAATAT 1184
|||||:|||||:
91 ArgLeuAspAsnGlySertYrAsnPheSertYrAspLysGlnAspAsnme 107
1185 TTCCCTTATTGACGAGAAAGCGAATGATGATCTACACAGCAACATCA 1234
|||||:|||||:
107 tSerPheGlyAspHisGlyThrHisValAlaGlyIleAlaIleAlaLysA 124
1235 ATCAAGGTGCTGGA.....GGATTATATTCCAGGAGCTTTTACGGTC 1278
|||||:|||||:
124 rGAspGlyAlaGlyIleHisGlyValAlaPheAspAlaSerIleGly 140
1279 TCGCGCTGAAT..... 1290
|||||:
141 ThrLysLeuAsnAspTyrGlyAsnArgAsnGlyArgGluGluLeuIleGI 157
1291 .....AACGAACCTTGGCAAGCGCGCGCTTCATA 1321

```



```

3779 AAAAACCCCTGGAGACAGCGATCGGCAACTGGGACGGGCTTGCCACAGCG 3828
827 .....SerLeuaspLylGlyTyrGlnSerAspAlaHis..... 837
3829 GCGGTTTTCGGGCAATACGGCATCGACAGGTTTCATACATCGGCATACAGCG 3878
838 .....SerAspAsnTyrHisLeuGlnLeuTyrGly 847
3879 GGGCGCGGGTTTATGACAGCGCGACCTTTGACAGCGCATGGAGCAAAA 3928
847 YAspLysArgPhe.....GlyAlaLeuAlaLeuArgAlaGlyLylThr 862
3929 TC.....CGCGCGCGCGTGGCATTCACGGCATTCAG 3960
862 YThrTrpHisArgIleAspThrSerArgSerValAsnTyrGlyAlaGln 878
3961 .....GCACATACCGCGCGGTTTGGCGGATTCGGCAT 3995
879 SerAspArgGlnLysAlaLysTyrAsnAlaArgThrGlyGlnLeuPheIle 895
3996 CGAACCGCGACATCGCGGCGACACGGCGTATTTCGTCCAAAAAGCGATTAC 4045
895 eGlu.....SerGlyTyrAspTrp 902
4046 GCTACGAAACGTCGATATCGCCACC...CCGGCGCTTTCATTCGCAACGCC 4092
902 hSerAspAlaValAsnLeuGlnProPheAlaAsnLeuAlaTyrThrHis 918
4093 TACCG..... 4098
919 TyrArgAsnGlnGluIleAsnGlnGlnGlyAlaAlaAlaLeuArgGln 935
4099 .....CGGGGCAATTAGCGAG 4114
935 YAspLysGlnSerGlnSerAlaThrAlaSerThrLeuGlnLeuArgAla 952
4115 ATTATTCATTCGAAACCGCG..... 4134
952 sPThrGlnTrpGlnThrAspSerValAlaIleAlaLeuArgGlyGlnLeu 968
4135 .....CAACAC.....ATTTCATTCAGCGCTTATTGGACCT 4166
969 GlnTyrGlnHisGlnTyrGlyLysLeuGlnArgLysThrGlnLeuMetPh 985
4167 GTCCCATACGATGCGCCTTCGGGCAAAAGTCGGAACCGGTCGTAATACCG 4216
985 eLysArgThrAspAlaIleAsp.....ValAsnSerV 997
4217 CCGATTGGCGTCAGGATTCGGCAAAACCGCGAGTGGCGAATGGCGGTA 4266
997 alProValSerArgAsp...GlyAlaIleLeuLysAla.....GlyVal 1010
4267 AACCGCGCAATC...AAGGTTTCACGCTGCTCCCTCAGCGCTGCCCGCG 4313
1011 AspValSerIleAsnLysAlaValAlaValLeuSerLeu.....Gln 1023
4314 CAAGGCGCGCACTGGAAGCGCAACAC...AGCGGGCGCATCAAAATTG 4360
1023 YTyrGlyGlyGlnLeuSerSerAsnHisGlnAspAsnSerValAsnAlaG 1040
4361 GGTACCGCGTGG 4371
1040 lYeuThrTrp 1043

```

```

seq_name: SwissProt_40:YCY3_HAUSQ
seq_documentation_block:
ID      YCY3_HAUSQ      STANDARD;      PRT;      437 AA.
AC      P21561;
DT      01-MAY-1991 (Rel. 18, Created)
DT      01-MAY-1991 (Rel. 18, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical 50.6 kDa protein in the 5' region of GYRA and GYRB (ORF

```


385 sar9.....glyargleuargvalarglys 394
4057 GTCAATATCGCACCCCGGCTTGATTCACCGCTACCGCGCGGCAT 4106
394 erThrAlaIaValProArg.....ProleuProArgInHis 406
seq_name: Swissprot_40:VG37_BPT2

seq_documentation_block:
ID VG37_BPT2 STANDARD: PRT: 1341 AA.
AC P07067;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Long tail fiber protein p37 (Protein Gp37) (Receptor recognizing
protein).
GN 37.
OS Bacteriophage T2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like phages
OX NCBI_TaxID=10664;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87112716; PubMed=3806672;
RA Riede I., Drexler K., Eschbach M.-U., Henning U.;
RT "DNA sequence of the tail fiber genes 37, encoding the receptor
recognizing part of the fiber, of bacteriophages T2 and K3".
RL J. Mol. Biol. 191:255-266(1986).
CC -1- FUNCTION: STRUCTURAL COMPONENT OF THE DISTAL-HALF TAIL FIBER.
CC -1- SUBUNIT: THE DISTAL HALF-FIBER CONTAINS TWO MOLECULES EACH OF GP36
CC AND GP37 AND ONE MOLECULE OF GP35.
CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: X04442; CAA28038.1; -
CC PIR: S09579; S09579.
CC Fiber protein.
CC
CC SEQUENCE 1341 AA; 143152 MW; 1B7981E7E5FA7F5 CRC64;

alignment_scores:
Quality: 205.50 Length: 1576
Ratio: 0.317 Gaps: 83
Percent Similarity: 41.117 Percent Identity: 18.718

alignment_block:
US-09-303-518D-649 x VG37_BPT2 ..

Align seg 1/1 to: VG37_BPT2 from: 1 to: 1341

181 AAAGCAAGTTGACAGTGGCGGCGCAAGATATTAGGTTTACAACAA.. 228
26 GtuglygluLeuAlaIleasnleuLysaspargThrIlePheThrLysas 42
229AAAGGAGTGTGTGGCAATATGACAAAGCCCGCATGTTG 274
42 pasperserLysanIleleaspleuserIleSerAlaIleGlyAsnIles 59
275 ATTTTCTGTGTG.....TCGCGTAAGCGCGTGGCGCA 309
59 erglyAsnIleThrIleThrLysPtyrThrGlnThrGlnLysPheasn 75
310 TTGTGGCGCATATATATTGTGAGCGGTGGACATACGCGCGCATATA 359
76 leuileglyProGlnIleValAlaser.....GlyIlytyr... 87

360 CAACGTTGATTTGGTGGGAGAGAGAGAAATCCCGATCAACATCGTTTA 409
88IleGluPhe..... 90
410 CTTATAAATTTGAAACGGAATTAATTATAAGCGAGCATTAAGCCAT 459
91AsnIlytyrArg.....ThrThrGlySer 97
460 CTTATGCGCGCGGATTCATATATGCGGTTTGATTAATTTGTC..... 504
98 GlyAlaIlePserGlyGlnHisThrAlaLysAlaProIlePheValaspLe 114
505ACAGATGCAAGACCTGTTAAATGACCAAGTT 535
114 userSerAlaIlePserThrSerGluTyrAsnProIleIleLysGlnArgp 131
536 ATATGATGCGCGG..... 549
131 helYaspspIlyThrPheSerleuGlyThrIleValSerGluGlySerleu 147
550AATATATGATCAAAATATATACCTGACCGTGTGTATTTGG 593
148 LysIleHisTyrIleAsnGluSer..... 155
594 GGCAGCGAGGCAATATTTGCGATCTGATGAAGATGACCCCAATACCGCG 643
156 .GlyAspserIlytyrP..... 161
644 AAAGTTATATCATATATGCAAGTCGATTTCTTGCTGTGGTGCAT 693
161 161
694 ACCTTTCACAAATATGATGATGATGCGAGTGCACATGACTAGTATGTA 743
162 ThrIleHisTyrIleAsnGluSer.....GlyIlyPheThrValAspIleGly..... 174
744 AAAATATTAACATATGATGATGATGATGATGATGATGATGATGATG 793
175GlyleuGlyValserGlyGlySerIleT 184
794 GCGACAGTGC.....TCACCATGTTT 816
184 hrThrSerGlyAsnIleAlaIleAlaLeuGlyAsnIleThrSerProGlnIle 200
817ATCTATGATGC 827
201 AsnThrLysAsnIleIleLeuAspThrLysAlaPheGlyGlnTyrAspSe 217
828 CCAAAAGCAAAAGTGTATTAATTAATGCGGTATGCAAGCGGCAACCCCT 877
217 rGln.....SerleuValGlnIlytyrValIlytyrProG 227
878 ATATAGGAAAGCAATGCTTCCAGCTGTTGCTGTA..... 915
227 LyrThrGlyGluIleAsnGlyIleAsnTyrLeuArgLysValArgAlaLys 243
916GATTGCTTATGATGAATCTTTCGCGAGATACCCCTTACGT 959
244 SerGlyGlyThrIleTyrHisGluIleAlaSerAlaGlnThr..... 257
960 ATTTCAGCAACACAGCTCAAAATGGAATATCTTTTACGAGCAATATA 1009
258GlyYasnaspspGluIleSerTrpTrpThrGlyAsnI 270
1010 ATGCGACAGAAATATCAATGCGCAACATGACACATTTCTGCTTAAT 1059
270 hrleuThrThrLysLeuMetClyLeuArgAsnAspGlyAlaMetValIleu 286
1060 AGATTAATAAACACGACCGTTCAATGTTTAATGTTCTTTATTCAGAGAC 1109
287 Arg.....ArgSerleuAlaIleGlyThrIleThrAlaAspIleu... 299
1110 AGCAAGAGAACCTGTTATATCATCTGACGAGTGTGTCAACAGTTATGCAC 1159

```
300 .....:|||||:|
1160 CCAGACGATATATGGAGAA.....AsnThrAsnAsnIYrGlyS 306
306 ePrProthPrMetGlyGluArgTYrIleAlaLeuGlyAspAlaIaThr 322
1180 AATATTTCTTTATTGACGAGAGAAAAGCGAATTGATCTACACAGCA 1229
323 GlyLeuLysTYrIleYsgInGlyValIYrAspLeuVal.....GlyAs 337
1230 CATCAATCAAGCTGCTGGAGATTATATTCCAAGAGATTATTACGCT 1279
337 nTrpAsnSerValAla.....SerIleT 345
1280 CGCCTGAATATACGAAACTTGGCAGCGCGCGCTTCATATCAATGAA 1329
345 hrProAspSerPheArgSerThrArgIYsAlaLeuPheGlyArgSerGlu 361
1330 GACAGTACCGTTACTTGGAAAGTAAAGCGCGTGGCAACGACGCGCTG 1379
362 AspGlnGlyIYrThrTrpThrMetProGlyThrAsnAlaIaLeuLeu 378
1380 C.....AAATCGCAAAAGCAGCGCTGCAGC 1405
378 rValGlnThrGlnIaAspValaAsnAsnIaGlyAspGlyGlnThrHis 395
1406 TTCAAAGCCAAAGGAGAAACCAAGCGCTGCATCAGCGTGGCAGGTACA 1455
395 Le.....GlyTYrAsnSerGly..... 400
1456 GTCATTTTGGATCAGCAGGACGACGATTAAGGCAAAAAACAACCTTAG 1505
401 .....GlyLysMetSerHisTYr.. 406
1506 TGAATCGCTTGTCGTACGCGCAGGCTACGCTGCACATGATCGCGTAA 1555
407 .....PheArgGlyLysGlyGlnThrAsnIleAsnThrGlnL 419
1556 AT.....CAGTTCACCCCGAC.....AACTCTATTGGCGCTTGGC 1593
419 ysgIYMetGluValaAsnProGlyIleLeuLysLeuValThrAspSerAs 435
1594 GCGCGACGTTTGATTTAAAGCGGCACTGCTGCTCCACCGATTCGA 1643
436 AsnValGlnPheTYrAlaAsnGlyThrValSerSerIleGlnArgIleY 452
1644 AATATCCGATGAAGGCGCGATGTTGCAACCAATCAAGAC..... 1686
452 S...PheAspAsnGlyLeuValLeuThrGlyAlaArgProAspGlyIleG 468
1687 ..AAGAATCCACCGTTACCATTCACGCAATTAAGATTTGCTACACG 1734
468 ILeuAspAlaProThrIaIaAspGlyThrIYsThrIleLeuTrpAla 484
1735 GGCATATACACAGCTTGATAGCAAAAAAATTGCTCTACACGTTG 1784
485 GlyGlyThrArgAlaGlyGlnAsnLysSerTYrValSerIleYsAlaTr 501
1785 GTTTGGCAGAAAGATACGACCAAAAGCAAGCGCGCTCAACCTTTT 1834
501 pGlyAsnSerPheAsnAlaSerGlyAspArgAlaArg...GluThrVal 517
1835 ACCAGCCGCGC.....GCAGAGACCGACCGCTG 1863
517 hGluValaGlyAspGlyGlnGlyPheHisPheTYrSerGlnArgValaIa 533
1864 CTGCTTTCGCGGAGACA.....AATTAAACGGCAA 1895
534 ProAlaProGlySerThrValGlyProIleGlnLeuArgValaAsnGlyG 550
1896 CATCAGCGCAAAAGCAAGCTGTTTTCAGCGCAGACCAACACCGCG 1945
```

```
550 yLeuLeu...ThrAlaGlySerIleValAlaSerGlySerIleThrThrg 566
1946 ACGCCATCAATCATTTAAACGACCATTTGGTCCGCAAAAAGAGGCATTC 1995
566 IuSerSerLeuAsnValaAsnAsnGlyLeuSer..... 576
1996 CGCGGGGAATGCTGTGGACAAACGATGATCAACCGCACATTTAAAGC 2045
576 ..... 576
2046 GGAATCTTCATTAATAA...GGCGGACAGCGGCTGTTCCGCAATG 2092
577 ValAsnGlyGlnAlaIaIaSerPheGlyGlyThrAla.....AsnA 589
2093 TTGCGCAAGTGAAGGCGATTTGGCATTTGACATACGCCCAACGACGTT 2142
589 IalaAsnArgIle.....TrpAsn.....AlaGlu 596
2143 TTTGCTGTGCGACCGCATCAAGCCACACATCTGACATCTTGGACGTC 2192
597 TYrGlyVal.....IlePheArgArgSerGlu.. 605
2193 GACGGGTCTGCATAATTGTGTGCAAAAAACATTACCGAGATTAAGTGA 2242
606 .....SerAsnPheTYrIleI 611
2243 TTGCTTCATTTAGTACAGACCGACATCAAGCGGCATGTCATCTTGGCGAT 2292
611 IeProThrAsnGlnAsnGluGlyIuSerGlyAspIle..... 623
2293 CAGGCTCATTTAATCTCACAGGCGCTGCCACACTCAACGCAATCTTAG 2342
624 HisSerSerLeuArgProValArgIleGlyLeuAsnAsnSpGlyAlaValG 640
2343 TGCATAATGCGCATACACGCTTATACAGTCACGACCAACAGCC.....ACC 2386
640 yLeuGlyArgAspSerPheIleValAspGlnAsnAlaLeuThrTrpI 657
2387 AAAACGGCACCTTAGCTCTGCGGCAT.....GCCCA 2421
657 IeAsnSerAsnSerArgIleAsnAlaAsnPheArgMetGlnLeuGlyGln 673
2422 GCACATTTAATCAAGCCCATTTAAAGCGCAACATCGGCTTGGCGCA 2471
674 SerThrTYrIleAspAlaGlyCysThrAspThrValArgProAlaGlyAl 690
2472 TGCTCATTTAATCTAAGGACACGCGCTGAACAGCGACGTGACGC 2521
690 agLYserPhe.....ValSerGlnAsnAsnGlu..... 699
2522 TTTCCGCGCACGCTAAGCA.....AACGTAAAGCTATTCGCA 2559
700 .....AsnValArgAlaProPheTYrMetAsnIleAsnArgThrAsp 713
2560 CTCACAGCTAATGCTCTCCCTACCGCAATGAAGCGATTCATTTGAAG 2609
714 ThrSerThrTYrValIProIleLeuLysGlnArgTYrValGlnGlyAsnSe 730
2610 CAGCGCGTTTACGCGCAAAATCAAGCGCGCAAGATPACGGCATTTCACT 2659
730 rCysTYrSerLeuGlyThrLeuIleSerThrGlyAspPheArgIleHisT 747
2660 TA..... 2661
747 yHisGlnGlyGlyAspAsnGlySerThrGlyProGlnLysAlaAspLeu 763
2662 .....AAAGCAGCAATGACAGCTGCGCTCAGCGCACGCA 2696
764 AlaTrpGlnPheArgArgAspGlySerPheArgSerProAsnLysIleG 780
2697 ATTA.....GCCAATTA..... 2709
780 uIleAsnAlaValThrIleGlyThrAspGlyAsnIleThrGlyGlyThrg 797
```

2710 ..AACCTTACAAAGCGCACCACTTAACTCAATCAATTCGGCTATGCGCCACGAT 2755
 797 LysAsnHelaLysAsnLeuAsnThrThrLeuAsn.....ArgLysThr 810
 2758 GCGGACGAGGCGCCAAACCGGACGCGTGCAGACAGATGGCGCGCGCCGCGCTTC 2807
 811 ThrValGlyGlyTrpAlaGlyIysSerSerValValGlyTrpLys..... 825
 2808 GCGCGCTTCGCGCGCTCCCTATTATCCGTTACACCGGCACACTTCGGTAG 2857
 826PheAlaThrValThrIleProGlnSer..... 834
 2858 AATCCCTTTTCAACACGCTGACGGGAAACGGCAATTGACGCTACAGGA 2907
 835ThioLysThrValThrPheLysIleSerGlyGlyAla 846
 2908 ACAATCCGCTTTATGTCGGAACTCTTCGGCTACGCGACGCGCAATTGAA 2957
 847 GlyPheAsnPheLysSer.....TyrSngAlaIle 857
 2958 GCTGCGCGGAAAGTTCCGAGAGCACTTACACTTGGCGGTCAACATACCG 3007
 857 rIleAlaGlu.....IleValIleArgThrG 866
 3008 GCACAGCAACT.....GCA 3021
 866 LysAsnAsnProLysGlyIleAsnAlaValLeuTrpAsnArgSerAspLeu 882
 3022 AGCCTGCACAAATTGACGGTAGAGGGAAGGAAAGACACAA..... 3063
 883 SerPheAsnGlnIleAlaThrMetAsnThrSerAspAspThrLysAspAsp 899
 3064CCGC 3067
 899 LysValPheCysGluGlyTyrThrAsnAlaLeuIleValGlyTyrSer 916
 3068 TGTCCCAAAACCTTATTACCCCTGCAAAACGACGCTCATGCGCGC 3117
 916 YsserGlnsSerSerValThrValGly.....LeuAsnGlyGly 930
 3118 GCGTGGCCTTACCAACTCATTCGCAAAAGACGCGAGTTCCGCTCATTA 3167
 931 Ile..... 931
 3168 TCCGTCAAAGACAGACAGCTTCCGACAACTC.....GGCAAGCGAG 3211
 932GlnProValValAspAlaLeuProGluGlnHisVal 944
 3212 AAGCCAAAAACAGCGCGGAAAAAGACAGACGCGCAAGCCTTGACGCGCTG 3261
 944 aLgLyThrValArgLeuLeuAsnAsn.....IleGlyGlyThr 957
 3262 ATTGCGCGCGCGCGC...GATCCGTGMAAAAGACAGAAAGCCTTGCGGA 3308
 958 PheAlaIleGlyGlnSerAspIleValThrArgGlyGlyIuTrValAlaAs 974
 3309 ACCGCGCGCG.....CAGGACGCGCGGGAATGTCGCG... 3342
 974 PAsnGlnLysGlyMetArgIleLysSerAsnGlyAsnAsnIleGlySerA 991
 3343ATTATGACGCGCGAGGAGACAGAAAAACGCGTGCAGCGC... 3381
 991 snAlaIleAlaIleLeuAlaGlnAsnAspGlyLysPheTrpIleLeuAlaThr 1007
 3382 GATAAAGACACCGCCTTGGCGGAACAAGCGGAAGCGGAAACCGCGCGCGC 3431
 1008 AspLysAsnThrAla..... 1012
 3432 TACACCGCGCTTCCCGCGCGCGCGCGCGCGCGCGGATTTGCCGCAAC 3481
 1013GluLysSerAspAlaAlaAsnGlyAspTrpAsnGlyL 1025

3542 GGCACACCCCAACCCGACCCCAACCGGAGCGGACCTGATCAGCCGTAT 3551
 1025 euArgPro.....Phe 1028
 3532 GCC.....AATAGCGTTTGAGTGA 3551
 1029 AlallelnmetalaspGlyatGValGlyMetasnhIsGlyLeuAsn11 1045
 3552 ATTTTCGCCACCGCTCAAC.....AGCGTTTCGGCGGAC 3566
 1045 ethGlyglGlyLeuAsnValThrGlyLeuAsnValThrSerGlyA 1062
 3587 AGCAGCAATTAAGACCGCGTATTTGCCGAAGACCGCGCAACCGCGTTTG 3636
 1062 snThrSerLeuGlyAsn11.....SerSerArgValValAlaArgTrp 1076
 3637ACAAGCGCATTCGGGACACCAACACACTACCTGCGCAAGATT 3680
 1077 ArgGlyAlaSerGlyTrpAlaAspAsnSerAspThrMetGlySerLys11 1093
 3681 CCGCGCGCTACCGCAACAAACCGACCTCGCCAA..... 3714
 1093 ethrPheMetAlaAspHisGlyAspLeuSerAsnSerAspSerTyrTrp 1110
 3715ATCGGTATGCAGAAAAACCTCGCGACGCGCGCGTGGCATCTG 3759
 1110 rolleValGlyAlaTyrSerAsnTyrGlySerAlaGlyTrp..... 1123
 3760 TTTTCGCACAAACGGAGCGAAACACTGTGCAGACGCGCATCGGCAATC 3809
 1124ArgLeuThrPheLeuIleuGly..... 1130
 3810 GGCACGGCTTGCCCAACGGCGCGCTTTTCGGGCAATACGCATGACAGGT 3859
 1130 1130
 3860 TCTACATCGCATCAGCGCGCGCGCGGTTTACGACGCGCACTT... 3906
 1131 ..TrpValGlySerGlyThrThrAlaGlyTyrTrpArgAspGlyIleuArg 1146
 3907TCAGCGGCATCGAGGCAAAATTCGCGCGCGGTGCTGCATTA 3950
 1147 IleArgGlyAspAsnAlaAsnGlyIn..... 1155
 3951 CGGCATTACGACAGCATACCGC..... 3972
 1156GlnAlaArgTrpArgPheThrMetAspGlyThrLeuAspCysP 1170
 3972 3972
 1170 roGlyLysValLeuLeuProGlnThrGlyAlaPheGlyAlaSnThrSer 1186
 3973 GCCGTTTCGGCGGA.....TTTCGC..... 3993
 1187 AsnGlyLeuGlyLysAsnSerIleThrPheGlyAspSerAspThrGly11 1203
 3994ATCGAACCGGACATCGCGCAACGGGCT 4021
 1203 eLysGlnAsnGlyAspGlyLeuLeuAspIleTyrAlaAsnSerValGln 1220
 4022 ATTTC...GTCCAAAAAGCGGAT...TACCGCTACGAAAACGTCAATATC 4065
 1220 alPheArgPheGlnAsnGlyAsnGlyLeuTyrSerTyrLysAsn11LeuSn... 1235
 4066 GCCACCCCGCGGCTTCGATTCACACGCGTAC...CGCGCGGCGATTAAAGC 4112
 1236AlaProAsnValTyrIleArgSerAspIleArgLe 1247
 4113 AGATTATCATTCAAACCGGCGCAAC 4140
 1247 uLysSerAsnPheLysProIleGluAsn 1256


```
623 ...SerProhe.....IleAlaGlySerGlnAlaLeuLeuLeu 634
790 .....TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAA 832
634 sGlnAlaLeuAsnAsnLysAsnSnpProPheTyrAlaTyrTyrLysGlnL 651
833 AGCAAAAGTGGTTAATTAAATGGGATATTCGAACCGCAACCCCTAATTA 882
651 eulysGlyThrAlaLeuThrAspPheLeuLysThr..... 662
883 GGAAGAAGCAATGGCTTCACAGTGGTTCGTAAGATGGTTCATATGATGA 932
663 ValGluMetAsnThrAlaGlnProLeu...AsnAspIleAsnTyr.... 676
933 AATCTTGGCTGGATACCATTCAGTATCTACGAACCAAGCTGCAAAATG 982
677 .....AsnAsnValIleValSerProTyrArgLysGlnG 687
983 GG.....AAATCTCTTTTAAAGCAGATTAATTAATGCG 1014
687 LylAlaGlyLeuValAspValLysAlaIleAspAlaLeuGluLysAsn 703
1015 ACAGGAAAAATCAATGCCAATCATGAACAAATCTCTGCTAATGATATT 1064
704 ProSerThrValValAlaGluAsnGlyTyrProAlaValAlaLeuLysAs 720
1065 AAAAAGCAGAACCGTTCAATTGTTTAAATGTTCTTTA..... 1101
720 PpherThrSerThrAspLysThrPheLysLeuThrPheThrAsnArgThr 737
1102 .....TCCGAGACAGCAGACAGACAGCTGTT 1125
737 hrHISGluLeuThrTyrGluMetAspSerAsnThrAspThrAsnAlaVal 753
1126 TATCATCTCGACAGTGGTGTCAACAGT.....TATCGACCCGAGCT 1166
754 TyrThrSerAlaThrAspProAsnSerGlyValLeuTyrAspLysLysII 770
1167 GAATTAAT.....GGAGAAAATATTTCTCTTATTAAGCAGAG 1201
770 eaSpGlyAlaAlaIleLysAlaGlySerAsnIleThr...ValProAlaG 786
1202 GAAAAGGC.....GAATTCATCTTACCGACGCAACATCAAT 1236
786 LylYsrThrAlaGlnIleGluPheThrLeuSerLeuProLysSerPheAsp 802
1237 CAAGGT.....GCTGAGATTTATTTCCAAAGAGATTTACGCT 1277
803 GlnGlnGlnPheValGlnLysPheLeuSnpPheLysGlySer..... 816
1278 CTCGCTGAATAATTAACGAATTTGGCAAGCGCGGCGCTCATATCACT 1326
817 .....AspGlySerArgLeuSnpLeuProT 825
1327 .....GAAGACAGTACCGTTACTTGGAAAGTA 1353
825 yMetGlyPhePheGlyAspTrpAsnAspGlyLysIleValAspSerLeu 841
1354 AAGCGCGTGGCAAAACGACCGCTGTCCAAATCGCAAA.....GGCAC 1397
842 AsnGlyIleThr.....TyrSerProAlaGlyLysPheGlyThr 855
1398 GCTGACGTTCAAGCAGCAAGGGGAAACCAAGGCTCG..... 1434
855 rValProLeu...LeuLysAsnLysAsnThrGlyThrGlnTyrTyrGly 871
1435 .....ATCAGCGTGGCGACGAGTCACTCATTTGGATCAGCAGCA... 1476
871 LylMetValThrAspAlaAspGlyAsnLysThrValAspAspGlnAlaIle 887
1477 .....GACGATAAAGC..... 1488
888 AlaPheSerSerAspLysAsnAlaLeuTyrAsnAspIleSerMetLysTyr 904

1489 .....AAAAACAAGCCTTTAGTGAATCGGCTTGTCAGCGCA 1528
904 rTyrIleLeuAlaArgAsnIleSerAsnValGlnValAspIleLeuSnpGly 921
1529 GGGGTACGGTGCACACATGATGCGATTAATGAGTTCAACCCGACAACTC 1578
921 LncLysAsnLysValThrThrLeuSerSerThrAsnArgLysLysThr 937
1579 TATTCGGCTTTCGGCGGAGCGTTTGATTTAAACGGCATTCGCTTTC 1628
938 TyrTyr.....AsnAlaHisSerGlnG 945
1629 GTTC..... 1632
945 nTyrIleTyrTyrAsnAlaProAlaTPrAspGlyThrTyrAspGlnA 962
1633 .....CACCGTATTCGAATACCGATGAAGCGCGATGATT..... 1668
962 rGAspGlyAsnIleLysThrAlaAspAspGlySerTyrThrTyrArgIle 978
1669 .....GTCAACCAACATCAAGACAAAGATCCACCGTTACATTACAG 1712
979 SerGlyValProGluGlyGlyAspLysArgGlnValPheAspAlaProph 995
1713 CAATTAAGATATT...GCTACAAACCGCATTAACACACAGCTGATAGCA 1759
995 eLysLeuAspSerLysAlaProThrValAlaGlnLysAlaLeuSerAlaL 1802
1760 AAAAAGAATTTGCTACACCGTTGGTT..... 1788
1802 ysrThrGluAsnGlyLysThrGlnTyrTyrLeuThrAlaGluAlaLysAsp 1808
1789 .....GGCGAAGAAAGATACGACCAAAAGC.....AACGG 1817
1809 AspLeuSerGlyLeuSnpAlaThrLysSerValLysThrGluIleAsnG 1845
1818 GCGGCTCAACCTT..... 1830
1845 uValThrAsnLeuAspAlaThrPheThrAspAlaGlyThrThrAlaAspG 1862
1831 .....GTTTACAGCCCGCGCGAGAGACCGCCTGCTGCTT 1869
1862 LylTyrThrLysIleGluThrProLeuSerAspGluGlnAlaLeu 1878
1870 TCCGCGGGAACAATTTAAACGCAACATCAACGCAAAACGCAACT 1919
1879 GlyAsnGlyAspAsn.....SerAlaGluLe 1987
1920 GTTTTCACGCGGACGACCAACGCGCAGCGCTACATATTAAACGACC 1969
1887 .....ThrAspAsnAlaSerAsnAlaThrAspGlnA 1100
1970 ATTTGCTCG...CAAAAGAGGCGATTCCTCGCGGGAATGCTGGGAC 2016
1100 sPalSerValGlnLysProGlySerThrSerPheAspLeuIleValAsn 1116
2017 AAC.....GACTGATCAACCGCACATTTAAAGCGGAAACTT 2054
1117 GlyGlyGlyIleProAspLysIleSerSerThrThrThrGlyArgIle 1133
2055 CCAAAATTAAAGCGGACAGCGGCTGTTCGCCGAATGTTGCCAAAGTGA 2104
1133 aAsnThrGlnGlyLys..... 1138
2105 AAGCGATTTGGCATTTGAGC...AATACCCCAACAGAGTTTGGTGTGC 2151
1139 ...GlyThrTyrThrPheSerGlyThrTyrProAlaIleValAspGlyThr 1154
2152 .....GCACCGCATCAAGCCACACATCTGACACGTTGGAGCTG 2192
1155 TyrThrAspAlaGlnGlyLysLysHisAspLeuAsnThrThrTyrAsp... 1170
```

2193 GACGGTCTGCAAAATGTGTGCAAAAAACCATTAACGAGTAAGTGA 2242
::: |||||
1171 ...AlaIatThrAsnSerPheThrAlaSerMet..... 1180
2243 TTGCTTACATGACAAAGCAGCATGCGGCAATGTCATCT...GCC 2289
::: |||||
1181 ...ProValThrAsnAlaAspTyrAlaAlaGlnValAspLeuTyrAla 1195
2290 GATCAGCGTCATTTAAATCTCACAGGCGTTGCCACATCAACGCGCATCT 2339
||| |||||
1196 AspTyrAlaHisThrGlnLeuLeuLeuHisPheAspThrLysValArgLe 1212
2340 TAGTGCAT...GCGATACAGTTATACAGTACGACCAAGCCCA 2383
||| |||||
1212 UMetAlaProThrPheThrAspLeuLysPheAsnAsnGlySerAspLnt 1229
2384 CCCAAACGCGCAACCTTAGCCTGCGCAATGCCCAAGCAACATTTAAT 2433
||| |||||
1229 hSerGlnAlaThrIleLysValThrGlyThrValSerAla.... Asp 1243
2434 CAAGCCACATTAAC...GCCAACATCGGCTCGGCAATGCTCAT 2480
||| |||||
1244 ThrLysThrValAsnValGlyHisThrValAlaAlaLeuAspAlaGlnH 1260
2481 TATCTAAGC...GACCAAGCGCTACAAACGCGAGTCTGACGCTTCCG 2527
::: |||||
1260 hHisPheSerValAspValProValAsnTyrGlyAspAsnThrIle.... 1275
2528 GCAACGCTAAGCAACGTAACCATTCGCGACATCAAGGTAAGTCTCC 2577
::: |||||
1276 ...LysValThrAlaThrAspLysAspLysAspLysAspLysThrThr 1287
2578 CTAGCCGATAGCAGCTA..... 2595
1288 ThrGlnGlnLysThrIleThrSerSerTyrAspProAspMetLeuLysLy 1304
::: |||||
2596 ...TTCCAT 2602
1304 sSerValThrPheAspGlnGlyValLysPheGlyThrAsnLysPheAsn 1321
2603 TTGAAACAGCGCTT.....ACC 2622
||| |||||
1321 LathSerAlaLysPheTyrAspProLysThrGlyLeuAlaThrIleThr 1337
2623 GCAAAATCAGCGCGGCAAGATACG.....GCAT 2654
||| |||||
1338 GlyValLysHisProThrThrThrLeuGlnValAspGlyLysGlnH 1354
2655 ACACCTTAAAGACAGC.....GAATGAGCGTGCCTGACGACGCG 2695
::: |||||
1354 eProLysAspAspLeuThrPheSerPheThrLeuAspLeuGlyThr. 1370
2696 AATTAGC..... 2703
||| |||||
1371 ...LeuGlyGlnLysProPheGlyValValGlyAspThrThrGlnAsn 1386
2704AA 2705
1387 LysThrPheGlnGlnAlaLeuSerPheIleLeuAspAlaValAlaProth 1403
2706 TTAAACCTTGACACGCGACCATTAACATTCGCGCTATCGCCACG 2755
||| |||||
1403 IleuSerLeuAspSerSerThrAspAla....ProValTyrThrAsn 1418
2756 ATGCGGCAAGGCGCAACCGGACGTGCGAAGATGCGCGCGCGCGCT 2805
||| |||||
1418 sProAsnPheGlnIleThrGlyThrAlaThrAspAsnAlaGlnTyrLeu 1434
2806 TCGCGCGCTGCGCGCGCTCCCTATTA..... 2832
||| |||||
1435 SerLeuSerIleAsnGlySerSerValAlaSerGlnTyrGlnAspLeas 1451
2832 2832

1451 nIleAsnSerGlyLysProGlyHisMetAlaIleAspGlnProValLysL 1468
2832 2832
1468 euleuGlnGlyLysAsnValLeuThrValAlaValThrAspSerGlnAsp 1484
2832 2832
1485 AsnThrThrThrLysAsnIleThrValTyrTyrGlnProLysLysThrIle 1501
2833TCCGTTACACCGCCCACTTCGTGAGAAATCCGTTTCACA 2872
::: |||||
1501 uAlaAlaProThrValThrProSerThrThrGlnProAlaGlnThrValT 1518
2873 CGCTGACGCTAAGCGCAATGTGACGCTGAG..... 2904
||| |||||
1518 hLeuThrAlaAsnAlaAlaAlaThrGlyGlnThrValGlnTyrSerAla 1534
2904 2904
1535 AspGlyGlyLysThrTyrGlnAspValProAlaAlaGlyValThrIleTh 1551
2905GAAACATTCGCTTATGTCG...GAACCTCGGCTACCGCA 2944
||| |||||
1551 rAlaAsnGlyThrPheLysPheLysSerThrAspLeuTyrGlyAsnGln 1568
2945 GCGACAAATGACGCTGCGGAAAGTTCCGAAGCAGCTTACACCTTGCG 2994
||| |||||
1568 eProAlaValAspTyr.....ValVal 1575
2995 GTCAACATATACGCGACAGCAACCTGCAAGCTCGAACATGAGGAGT 3044
||| |||||
1576 ThrAsnIleLysAlaAspAspProAla.....GlnLeuGlnAlaAl 1589
3045 GGAAGAAAGCAACCAACCGCTGTCGAAACCTTAATTTCACCTTC 3094
||| |||||
1589 a.LysGlnGlnLeuThrAsn.....LeuIleAlaSerAla 1600
3095 AAACGACACAGCTGATGCGCGCGCTGCGTTACCACTCATCGGCAAA 3144
||| |||||
1601 LysThr.....LeuSerAlaSerGlyLysTyrAspAspAlaThrTh 1614
3145 GACGCGAGTCCGCTGCATATCCGTCAAGAACAGACCTTTCGGA 3194
||| |||||
1614 rThrAlaLeuAlaAla..... 1619
3195 CAAACTCGCAAGCAGAAACCAAAACAGCGCGGCAAAAGCAACGCGC 3244
||| |||||
1620AlaThrGlnLysAlaGlnThrAlaLeuAspGlnThrAsn..A 1633
3245 AAAGCTTGACCGCTGATTTGCGCGCGCGCGATGCTCGAAAAAGCA 3294
||| |||||
1633 lAserValAspSerLeuThrGlyAlaAsnTyrAspLeuGlnThrAlaIle 1649
3295 GAAAGCTTGCGCAACCGCGCGCGCGAGCGGCGGAAAAATGTCGCGAT 3344
||| |||||
1650 AsnGlnLeuAla..... 1653
3345 TATCAGCGGAGGAGAAAGAAAAAGCGGTGCGAGGATTAAGACCG 3394
||| |||||
1654AlaLysLeuProAlaAspLysLysThs 1663
3395 CCTTGCGCAACAGCGGAAGCG..... 3417
||| |||||
1663 eThrLeuAsnGlnLeuGlnSerValLysAspAlaLeuGlyThrAspLeu 1679
3418GAACCGCGCGCTACG...ACGCGCTTCCCGCGCGCGCG 3458
||| |||||
1680 GlyAsnGlnThrAspProSerThrGlyLysThrPheThrAlaAla.... 1694
3459 CGCCGCGCGGATTTGCGCAACTGCAACCCCAACCCGACCCCAACCGC 3508
||| |||||

1695LeuAspAspLeuValAlaGlnAlaGlnAlaGlyThrng 1707
 3509 AGCGGACGATGATGACCGCTTATGCCAATAGCGGTTGAGTAATTTCC 3558
 1707 LthrAsp.....AspLndleuGln 1713
 3559 GCGACGCTACAGCGCTTTCCCGGTACAGAGCAATTAGACCGGCTATT 3608
 1714 AlaThrLeuAlaLysile.....LeuAspGluValLe 1724
 3609 TCCCGGAG.....ACCGCGGACACCGG..... 3631
 1724 uLaLysLeuAlaGlnGlyLeuLysAlaAlaThrProAlaGluValGly 1740
 3632TTTGGCAAGCGGATCCGG.....ACACC 3657
 1741 AsnAlaLysAspAlaAlaThrGlyLysThrTrpTyrAlaAspLeuAlaAs 1757
 3658 AAACACTACCGTTCCAGATTTCCGCGCTACCGCCACCAACACGACCT 3707
 1757 PThrLeuThrSerGlyGlnAlaSerAlaAspAlaSerAspLysLeuAlaH 1774
 3708 GCGCCAAATCGGTATGCAAAAACTCGGACGCGGCGGCGGATCC 3757
 1774 LsLeuGlnAlaLeuGlnSerLeuLysThrLysValAlaAlaAlaValGlu 1790
 3758 TGTTCGACACCGGACCGCAAAACCTTCGACGAGGATCGGCAAC 3807
 1791 AlaAlaLysThrValGlyLysGlyAspGlyThrThrGlyThrSerAs 1806
 3808 TCGG.....CACGCTTCCCGACGCGGCGGCTTCGCGCAATACGG 3848
 1806 PLYGlyGlyGlyGlyGlyThrProAlaProAlaProGlyAspThrG 1822
 3849 CATGACAGGTTCTACATCGGATCGGCGGCGGCGGTTTACGACGG 3898
 1822 LysLysAspLysGlyAspGlyLysSer..... 1830
 3899 GCAGCCTTCAGACGCGGACGCAAAATCCGCGCGCGCTGCTGCT 3948
 1831GlnProSerSerGlyGlyAsnLeuProThrLysProAla 1844
 3949 TAGGCGCTTACGACGATACCGCGCGGTTTCGCGGATTCGCGATCGA 3998
 1844 hThrThrSerThrThr.....ThrAspAsp 1852
 3999 ACCGACATGCGCGCAACGCGCTATTGCTCCAAAAGCGGATTACCG 4046
 1853 ThrThrAspLysGlnGlyLysLeuThrSerGlyLysGlyAlaLeuPro 1868
 seq_name: SwissProt_40:120K_RICRI
 seq_documentation_block:
 ID 120K_RICRI STANDARD; PRT; 1300 AA.
 AC P14914;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE 120 kDa surface-exposed protein.
 GN P120.
 OS Rickettsia rickettsii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiidae; Rickettsia.
 OX NCBI_TaxID=783;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-R.
 RX MEDLINE=90136087; PubMed=2515418;
 RA Gilmore R.D. Jr., Jostle N., McDonald G.A.;
 RT "Cloning, expression and sequence analysis of the gene encoding the
 RL 120 kD surface-exposed protein of Rickettsia rickettsii";
 Mol. Microbiol. 3:1579-1586(1989).
 CC -!- FUNCTION: MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS

CC RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 CC S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -!- MISCELLANEOUS: ITS C-TERMINUS POTENTIALLY MAY BEAR THE EPTIOPES
 CC CONFERRING ANTIGENICITY TO THE PROTEIN.
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC
 CC EMBL: X16353; CAA34402.1; -
 CC PIR: S07575; S07575.
 DR InterPro: IPR003858; rompa_rompb.
 DR Pfam: PF02708; rompa_rompb; 1.
 DR Antigen: Glycoprotein; Cell wall; S-layer.
 KW
 FT CARBOHYD 66 66
 FT CARBOHYD 86 86
 FT CARBOHYD 103 103
 FT CARBOHYD 147 147
 FT CARBOHYD 268 268
 FT CARBOHYD 330 330
 FT CARBOHYD 375 375
 FT CARBOHYD 415 415
 FT CARBOHYD 424 424
 FT CARBOHYD 430 430
 FT CARBOHYD 436 436
 FT CARBOHYD 444 444
 FT CARBOHYD 515 515
 FT CARBOHYD 547 547
 FT CARBOHYD 553 553
 FT CARBOHYD 593 593
 FT CARBOHYD 698 698
 FT CARBOHYD 710 710
 FT CARBOHYD 799 799
 FT CARBOHYD 800 800
 FT CARBOHYD 826 826
 FT CARBOHYD 844 844
 FT CARBOHYD 861 861
 FT CARBOHYD 879 879
 FT CARBOHYD 920 920
 FT CARBOHYD 926 926
 FT CARBOHYD 1116 1116
 FT CARBOHYD 1128 1128
 FT CARBOHYD 1140 1140
 FT CARBOHYD 1146 1146
 FT CARBOHYD 1211 1211
 SO SEQUENCE 1300 AA; 132801 MW; E09E52C3F647243D CRC64;
 alignment_scores:
 Quality: 203.50 Length: 1309
 Ratio: 0.344 Gaps: 55
 Percent Similarity: 45.225 Percent Identity: 19.099
 alignment_block:
 US-09-303-518d-649 x 120K_RICRI ..
 Align seg 1/1 to: 120K_RICRI from: 1 to: 1300
 1009 AATGCGACAGGAAAAATCAATGCAATGCAATGCAATGCTGCTGCTAA 1058
 |||||.....
 7 AsnAlaThrGlyGlnValAlaSerPheArgHis..... 16
 1059 TAGATTAAACACGAAACGTTCAATTGTTTATTTCGAGAGA 1108
 ..:..:..:..:
 17lleValAspVal..... 20

1109 CAGCAGAGAACTGTTATCATCTGCTGCTGCTCAACAGTTATCGA 1158
21 GlyAlaAspGlyThrThrAlaPheLys 29
1159 CCC AGACTGAATPANGAGAAAATATTCCTTTAT 1193
30 ThrAlaIaSerLysValThrIleThrGlnAspSerAsnPheGlyAsnTh 46
1194 TCACGAGGAAAA GCGCAATTCGATAC 1219
46 rAspPheGlyAsnLeuAlaIaGlnIleLysValProAsnAlaIleThrL 63
1220 TTACGACGAAACATC AATCAAGT GCTGGA 1248
63 eutThrLysAsnPheThrGlyAspAlaSerAsnProGlyAsnThrAlaGly 79
1249 GGATTATATTTCCAAAGAGATTTCAGGCTCGCTCGCTGAATAACGGAAC 1298
80 ValIleThrPheAspAlaAsnGlyThrLeuGlnSerAlaIaSerAlaAsp.. 95
1299 TTGCGAAGCGCGGCGCTCATATCAGTGAAGACAGTACCGTTACTTGA 1348
96 AlaAsnValAlaValThrAsnAsnIleThrAlaIle...G 108
1349 AAGTAAACGCGCTGCGCAACGACCGCTGCTCCAAATCGGCAAGGACG 1398
108 IuAlaSerGlyAlaGlyValIaGlnLeuSer GlyThr 120
1399 CTGCGACTTCACGCCAAGGGGAAACCAAGCGCTCGATC...ACCGTGG 1445
121 HisAlaIaGlnLeuAlaGlyLeuAlaGlySerIlePheLysLeuAl 137
1446 CGACGCTACAGTCATTTTGATCAGCAGCAGATTAAGCAAAAAC 1495
137 aAspGlyThrValIle AsnGlyLysVala 147
1496 AAGCTTTAGTGAATCGGCTGCTGACGGGACGAGGCGCAAGCTG 1545
147 snGlnThrAlaLeuValGlyGlyAlaLeuAlaIaGlyThrIleThrLeu 163
1546 AATCGCATATATCAGTCAACCCCGAC 1572
164 AspGlySerAlaThrIleThrGlyAspIleGlyAsnAlaGlyAlaIa 180
1572 1572
180 AlaLeuGlnAlaThrIleThrLeuAlaAsnAlaLysLysThrLeuThrL 197
1573 AACTCTATTTGCGCTTTCGCGGACGCTTGGATTA... 1611
197 euGlyGlyAlaAsnIleIleGlyAlaGlyGlyThrIleAspLeuGln 213
1612 ...AAGCGCATTCGCTTTCGCTTCCACCGTATTCAAAAT... 1647
214 AlaAsnGlyGlyThrIleLysLeuThrSerThrGlnAsnGlnLeuVala 230
1648 ACCGATGAGGCGCGATGATTCACAC 1675
230 IAspPheAspLeuAlaIleAlaThrAspGlnThrGlyValValaAspAla 247
1676 ACAATCAAGACAAAGAAATCACCGTTACCATTAAGCAATTAAGATAT 1725
247 eSerLeuThrAsnAlaGlnThrLeuThrIleAsnGlyLys...Ile 261
1726 GCTACACCGCAATACAC ACCTGATAG 1757
262 GlyThrIleGlyAlaAsnAsnLysThrLeuGlnPheAsnIleGlyLys 278
1758 CAAAAAAGAAATTCCTACACAGGTTGTTGGCAGAAAGAT 1800
278 rSerLysThrValLeuSerAsnGlyAsnValAlaIleAsnGlnLeuValI 295
1800 1800

295 IeGlyAsnAspGlyAlaValGlnPheAlaHisAspThrIleThrL 311
1801 ...ACGACCAAAACGAAACGGCGGCTCAACCTGTTACACGCGCGC 1847
312 ArgThrThrAsnAlaIaIaGlyGlnGlyLysIleIlePheAsnProVala 328
1848 AAGAGACCGCACCTGCTGCTTCGCGGAGACAAATTAACGGCAACA 1897
328 IAsnAsnGlyThr...ThrLeuAlaIaGlyThrAsnLeuGly... 341
1898 TCACGCAACAACGCGCAACTGTTTTCAGCGGACAGACCAACGCGCAC 1947
342 ...SerAlaThrAsn... 345
1948 GCCTCAATTCATTTAAAGCATTGTCGCCAAAAAGAGGCAATTCCTCG 1997
346 ProLe 347
1998 CCGGAAATTCGTGTGGACAAACGACTGATCAACCGCATTTAAACGG 2047
347 uAlaGlnIle 350
2048 AAACTTCCAAATTAAGCGGACGAGCGGCGGTTTCCGCAATGTTGCC 2097
351 ...AsnPheGlySerLysGlyAlaAsnValAspThrValLeuAsnValGly 366
2098 AAGGTAAAGCGGATTCGATTTGAGCAATCAACGCCCAAGGATTTTGG 2147
367 Gln G1 368
2148 TGTGCGACCGCATCAAGCCACACATCTGTACAGTTCGAGCTGG... 2193
368 yAlaIaLeuValAlaThrAsnIleThrThrAspIlaAsnValGlyS 385
2194 ACGGCTGACAAATTCGTGCAAAAACCAT... 2226
385 ePheValPheAsnAlaGlyGlyThrAsnIleValSerGlyThrValIle 401
2227 ACCGACGATTAAGT 2240
402 GlyGlnGlnLysAsnLysPheAsnThrValAlaLeuGlnAsnGlyThrTh 418
2241 GATTGCTTCATTCAGTAAGACGACATCAGCGCAATGTCGATCTGCG 2290
418 rValLysPheLeuGlyAsnAlaThrPheAsnGlnAsnThrThrIleAla 435
2291 ATCAGCGCTCATTTAAATCTCACAGGCGTCCACACTCAACGCAATCTT 2340
435 IAsnSerThrLeuGlnIleGlyGlyAsnThrThrAlaAspCysValaIa 451
2341 AGTCAATATGCGCATCACCTTATACAGTCACGCAACGCGCAACCA 2390
452 SerIlaAspGly...ThrGlyIleValGlnPheValAsnThrGlyProIle 468
2391 CCGCAACCTTAAGCTCGTGGGCAATGCCCAAGCAACATTTAAT..CAAGC 2439
468 hrValThrLeuAsnLysGlnAlaIaIaProValAsnAlaLeuLysGlnIle 484
2440 ACATTAAACGCAACACATCGGCTTCGCGCAATGCTTCATTTATCTAAG 2489
485 ThrValSerGlyProGlyAsnValValIleAsnGlnIleGlyAsnAlaG 501
2490 CGACGAC GCGCTCAAAACGCGCAGTC 2515
501 yAsnHisHisGlyAlaValThrAspThrIleAlaPheGlnAsnSerSerL 518
2516 TG 2517
518 euGlyAlaValAlaPheLeuProAlaGlyIleProPheAsnAspAlaGly 534
2518 ...ACGCTTCCGCAACGCTAAGCAACGTAAGCAATTCGCGACTCAA 2564
.....

535 AsnThrMetProLeuThrIleLeuSerThrValGlyAsnLysThrAlaIy 551
2565 CGGTATATGTCCTCCTAGCCGATAAGCAGTATTCATTTTGAAGCACC 2614
551 sglly...PheAspValProSerValValIleuGlyValAspSerValI 567
2615 GCTTACCGCGCAATATCAGCGC..... 2637
567 AlaIAspGlyValIleGlyAspGlnAsnAsnIleValGlyLeuGly 583
2638 ...GGCAAGATACGGCATATACCTTAACAAACAGCCGATGCGC 2684
584 LeuGlySerAspAsnGlyIleIleValAsnAlaThr.....ThrLeuTy 598
2685 GTCAGGCAACGGATTAGCAATTAACTTGACACAGCCGACCATTAAC 2734
598 rAlaGly.....IleSerThrLeuAsnAsnAsnGlnGlyThrValThrL 613
2735 TCAATTCGCTATCGCCAGATCGCGAGCGGCGCAACCGCGAGTGC 2784
613 euserGlyGlyVal...ProAsnThrProGlyThrValItyrGlyLeuGly 628
2785 ACAGATGCGCGCGCGCGCTTCGCGCTTCGCGCGCTTCCTATTATC 2834
629 ThrGlyIleGlyAlaSerLysPheLysGln..... 638
2835 CGTTACACCGCCCACTCGGTAGATCCGTTTCAACAGCCTGACGCTAA 2884
639 .ValThrPheThrThrAspTyrAsnAsnLeuGlyAsnIleIleAlaThrA 655
2885 AC..... 2886
655 snAlaThrIleAsnAspGlyValThrValThrThrGlyGlyIleAlaGly 671
2887GCCAATTGAAAGCTGAGGAGAAC 2909
672 IleGlyPheAspGlyLysIleThrLeuGlySerValAsnGlyAsnGlyAs 688
2910 ATTCCGCTTATGTCGAACTCTGCGGCTACCGCAACCGCAATTCGAAC 2959
688 nValArgPheAlaAspGlyIleLeuSerAsnSerThrSerMetIleGlyT 705
2960 TGGCGGAAAGTTCCGAGGCACTTACACCTTGCG.....GTC 2997
705 hrThrLysAlaAsnAsnGlyThrValThrTyLeuGlyAsnAlaPheAla 721
2998 AACATATCCGCAACGAA.....CCTGCAAGCCTCGAACAATGACGCT 3041
722 GlyAsnIleGlyAspSerAspThrProValAlaSerValArgPheThr.. 737
3042 ACTGGAGGAAAGAACAAACACCGCTTCGAAACCTTAATTTCACCC 3091
738GlySerAspSer.....GlyAlaGlyLeuIndLysAsnI 749
3092 TGCAAACGCAACAGTCGATGCGCGCGGCTTACCAATCATCGCC 3141
749 LeTySerGlnValIleAspPheGlyThrTyAsnLeuGlyIleValAsn 765
3142 AAAGCGCGCGATTCGCGCTCATATTCGGTCAAAAGACAGCTTTC 3191
766 Ser.....AsnIleIle..... 769
3192 CGACAACTGGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3241
770LeuGlyLysGlyThrThrAlaIleAsnGlyLysIleAspLeuV 784
3242 CGCAAGCCTTGACGCGCTGATTCGCGCGCGCGCATGCGCGCAAGAG 3291
784 alThrAsn...ThrLeuThrPheAlaSerLysThrSerThrProGlyAsn 799
3292 ACAGAAAGCCTTGCGCAACCGCGCGCGAGGCGGCAAAATGCTCG 3341
800 AsnThrSerIleGlyThrThrLeuThrLeuAlaAsnGly...AsnIleG 815
3342 C...ATTATGACGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3388
815 yHisIleValIleLeuGlnGlyAlaGlnValAsnThrThrThrThrGlyT 832
3389 ACACCGCCTTGCGCAAGCGCGAGCGGAGAGAGAGAGAGAGAGAGAG 3435
832 hrThrThrIleLysValGlnAspAsnAlaAsnAlaAsnPheSerGlyThr 848
3436 ACCGCGCTTCCCGCGCGCGCGCGCGCGCGCGGATTTGCCAGTACA 3485
849 GlnThrTyThrLeuIleGlnGlyGlyAlaArgPheAsnGlyThrLeu.G 865
3486 ACCCCCAACCGCAGCCCAACCGCAGCGGACCGATTCGCGCTTAAGCCA 3535
865 Ty.....SerProAsnPheAlaValThrGlySerAsnArgPhe... 877
3536 ATACGCGTTTATGATTAATTTTCGCGCGCGCTCAACAGCGTTTTCGCG 3585
878ValAsnTySerLeuIleArgAlaAlaAsnGlnAspTy 890
3586 C.....AGCAGATTAGACCGCGCTATTTCGCGAGACG.....G 3620
890 rValIleThrArgThrAsnAsnAlaGlnAsnValValThrAsnAspIleA 907
3621 CCGCAACCGCGTTTGCACAGCGGATCCGCGGACACCAACCACTACGCTT 3670
907 IaAsnSerProPheGlyGlyAlaProGlyValAspGlnAsnValThrThr 923
3671 CGCAAGATTTCCGCGCGCTACCGGCAACCAACCGACCTGCGCAATCGGT 3720
923 923
3721 ATGCAGAAAACCTCGCAGCGCGCGCTGCGCATCTGTTTTCGACAA 3770
923 923
3771 CCGGACGGAACACCTTCGAGAGCGGATCGGCACTCGCGACGCGCTG 3820
924PheValAsnAlaThrAsnThrAlaAlaTyAsnAsnLeu 936
3821 CCCACGCGCGCTTTCGCGCAATACGCGATCGAGACAGTTCACATCGCG 3870
937 LeuLeuAlaLysAsnSerAlaAsnSerAlaAsnPheValGlyAlaIleVa 953
3871 ATCAGCGCGCGCGGCTTTCAGCAGCGGACGCTTCAGAGCGCATCGG 3920
953 ThrAspThrSerAlaAlaIleThrAsnValGlnLeuAspLeuAlaLysA 970
3921 AGGCAAAATCCGCGCGCGCTGCTCATTCAGGCACTTCAGGCAAGATAC 3970
970 spIleGlnAlaGlnLeuGlyAsnArgLeuGlyAlaLeuAlaGlyTyLeuGly 986
3971 GCGCGCGGTTGCGCGGATTCGCGATCGAATCGGCAATCGCGCAACGCG 4020
987 ThrProGlnThrAlaGlnMetAlaGly...ProGlnAlaGlyAlaIleSer 1002
4021 TATTTGTCAAAAGAGGATTTACCGGCTACGAAACGTCATATCCGCAC 4070
1003 AlaAlaValAlaIleGlyAspGlnAlaIleAspAsnValAlaTyGlyI 1019
4071 CCGCGCGCTTCGATTC.....AACGCTACCGCGCGCGCATTA 4108
1019 eTrpAlaLysProPheTyThrAspAlaHisGlnSerLysLysGlyGlyL 1036
4109 AGCGAGATTTATTCATAACCGCG..... 4134
1036 euAlaGlyTyTyLysAlaLysThrThrGlyValIleGlyLeuAspThr 1052
4135CAACATATTCATCAGCGCTTATTTAGCGCTGCGCATAC 4175
1053 LeuAlaAsnAspAsnLeuMetIleGlyAlaAlaIleGlyIleThrLysTh 1069

```

4176 CGATGCCGCTTCGGGCAAACTCCGAACACGCCGTACACGGCGATTGG 4225
|||||
1069 rAsp.....IleYSH 1073
4226 CTCAGGATTCGGCAAAACCCGAGTCGGATGAGGGCGTAAACGCCGA 4275
|||||
1073 tsglnAspTylYsLys.....GlyAspLysThrAsp 1083
4276 ATCAAGGTTTCACGCTGTCCCTCCAGCGCCGCCCAAGGCCCGCA 4325
|||||
1084 ValasnGlyPheSerPheSerLeuTyrGlyAlaGlnLeuValLysAs 1100
4326 ACTGGAACGGCAACACACGCGC 4347
|||||
1100 nPhePheAlaGlnGlySerAla 1107

seq_name: SwissProt_40:PMP6_CHLPN

seq_documentation_block:
ID PMP6_CHLPN STANDARD; PRT; 1276 AA.
AC Q92899; Q9JRW2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pmp6 precursor (Polymorphic membrane
protein 6).
CN PMP6 OR CPN0444 OR CP0309.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiales.
OX NCBI_TaxID=83538;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gall S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uitterback T., Berry K., Bass S.,
RA Liner K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Moyn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shiba T., Ishii K., Hattori M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CHL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
[4]
RN SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
(POTENTIAL).
[5]
RN SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
-----
CC EMBL; AE001627; AAD18588.1; -
CC DR EMBL; AE002193; AAF38166.1; -

```

```

DR EMBL; AP002546; BAA98652.1; -
DR PHCI-2DPAGE; Q92899; -
DR TIGR; CP0309; -
DR InterPro; IPR003368; DUF145.
DR InterPro; IPR003357; OMP.
DR Pfam; PF02415; DUF145; 2.
DR Pfam; PF02385; OMP; 1.
KW Outer membrane; signal; Multigene family; Complete proteome.
FT SIGNAL 1 23
FT CHAIN 24 1276
FT CONFLICT 421 421
FT FT
FT FT
FT FT
SQ SEQUENCE 1276 AA; 132127 MW; C97FA062C88E0B CRC64;

```

```

alignment_scores:
Quality: 201.50 Length: 1072
Ratio: 0.411 Gaps: 58
Percent Similarity: 45.709 Percent Identity: 21.362

```

```

alignment_block:
US-09-303-518D-649 x PMP6_CHLPN ..

```

```

Align seg 1/1 to: PMP6_CHLPN from: 1 to: 1276

```

```

1570 GACAACTCTATTTCGGCTTCGGCGGACGTTGGATTAAACGGG.. 1617
|||||
75 AspLysSerCysPheThrAsnThrGlyGlyAlaLeuSerPheValGlyAl 91
1618 ...CATTCGCTTCGTTCCACCGCTATTCAAATAACGATGAAGGCGCA 1663
|||||
91 AspHisSerLeuValLeuGlnThrIleAlaLeuThrHisAspGlyAla 108
1664 TGATGTCACCAACAAT..... 1680
108 lAlaAsnThrAsnThrAsnThrAlaLeuSerPheSerGlyPheSerSerLeu 124
1680 ..... 1680
125 LeuIleAspSerAlaProAlaThrGlyThrSerGlyGlyAlaIle 141
1681 .....CAAGCAAAAGAAATCCACCGCTTACCTTACAGCAATTAAG 1720
|||||
141 eCysValThrAsnThrGlnGlyGlyThrAlaThrPheThrAspAlaAs 158
1721 ATATGCTCTACAAACCGGCAATTAACAACAGCTTGATAGCAAAAAGAAATT 1770
|||||
158 eValThrLeuGlnLysAsnThrSerGlyLysAspGlyAlaAlaValSer 174
1771 GCCTACAC.....GGTGGTTGGCGCA 1793
|||||
175 AlaTyrSerIleAspLeuAlaLysThrThrAlaAlaLeuLeuAspG 191
1794 GAAAGATACCAACCAAAACGAGCGGGGCTCAACCTGTTTACCGACCG 1843
|||||
191 AsnThrSerThrLysAsnGlyGly..... 199
1844 CCGCAAGAACCGCACCTCTGCTTCGCGCGGCAAAATTTAAACGCG 1893
|||||
200 .....AlaLeuCysSerThrAlaAsnThrThrValGlnGly 211
1894 AACATTCACGCAACCAACGCAAACTGTTTTCAGCGGCAACCAACCC 1943
|||||
212 Asn.....SerGlyThrValThr.. 217
1944 GCACGCCATCAATCAATTTAAACGACCACTTGCGCAAAAAGAGGATTC 1993
|||||
218 .....PheSerSerAsnThrAlaThrA 225
1994 CTCGGGGGAATGCTGTGGGCAACGATGATCAACCGCACATTTAA 2043
|||||

```

225 splysglyglylletyrslysglu.....Lys 235
2044 GCGGAAACTTCGAATTAAAGCGCAGCGGTGGTTCCCGCAATG 2093
:::|||||::: ||| ::::: |||:::
236 AspSerThrLeuaspAlaasnThrGlyValValThrPheIysSerAsnTh 252
|||::|::| ||| |||
2094 TGCCAAGTGAAGAAGCGCATGGCAT.....TTGAGCA 2125
::|::|::|::| |::|::|::|
252 rAlaIysThrGlyAlaTrpSerSeraspAsnLeuAlaLeuThrG 269
:::|||||::: ||| |||
2126 ATCAGCCCCAACACTTTT.....GGTGCGACCG 2157
:::|||||::: ||| |||
269 LysAsnThrGlnValLeuPheGlnGluAsnLysThrThrGlySerAlaAla 285
:::|||||::: ||| |||
2158 CATCAAAGCGAC.....ACAATCTGT..... 2178
:::|||||::: ||| |||
286 GlinalaAsnAsnProGlnGlyCysGlyAlaValIleCysCysteLeuAl 302
2179 .ACAGCTTGCGACTGAGCGGGCTG....ACAATGTGTGCAGAAAAA 2221
|||::|::| ||| ||| |||
302 aThrAlaThrAspLysThrGlyLeuAlaIleSerGlnAsnGlnGluMetS 319
2222 CCATTACGGAGATAAAGTATTGCTTCA.....TTGACT 2256
:::|||||::: ||| |||
319 eRpHeThrSerAsnThrThrAlaAsnGlyAlaIleTyralaThr 335
:::|||||::: ||| |||
2257 AAGACGCACATCACGCGCAATGTCGATCTGGC..... 2289
|||::|::| ||| |||
336 LysCysThrLeuAspGlyAsnThrThrLeuThrPheAspGlnAsnThrAl 352
2290GATCAGCGTC 2299
352 aThrAlaGlyCysGlyAlaIleTyrrThrGlnThrGlnAspPheSerL 369
2300 ATTTAATCTCACAGGGCTTGGCACACTCAAGCGCAATCTTAGTCAAT 2349
::: |||||::: |||::: |||:::
369 euLySGlySerThrGlyThrValThrPheSerThrAsnThrAlaLysThr 385
2350 GCGCATACAGCTTATACAGTCAAGCCACGCCAACCCAAAACGCAACT 2399
|||::: |||::: |||:::
386 GlyAlaValaLeuTyrrSer.....LysGlyAsnSe 395
2400 TAGCTCTGTGGCAATGCCCAAGCAACATTT.....AATCAGCCACAT 2443
|||::|::| |||::|::| |||
395 rSerLeuThrGlyAsnThrAsnLeuLeuPheSerGlyAsnLysAlaThrG 412
2444 TAAACGCAACACATCGGCTTGC.....GCCAATGCTTAATTT 2481
:::|||||::: |||::: |||:::
412 LyProSerAsnSerSerAlaAsnGlnGlnGlyCysGlyAlaIleLeu 428
2482 AATCTAAGGACCACGCC.....GT 2501
::: ||| |||
429 SerPheLeuGlnSerAlaSerValSerThrLysGlyLeuTrpIleGl 445
2502 ACAAAAGCGCACTGACGCTTTCGCGCAAGCGTAAGCGAAACGTAAGC. 2550
::|::|::: |||::|::| |||::|::|
445 uAspAsnGlnAsnValSerLeuSerGlyAsnThr..AlaThrValSerG 461
2551CATTCGCACTCAACGTAATGTCTCCCTA 2580
|||::|::| |||::|::| |||::|::|
461 lYgLYAlaIleTyrrAlaThrLysCysAlaLeuHisGlyAsnThrIleu 477
2581 GCCGATAAGCAGTATTCATTTTGAAGAAGCGGCTTACCGGCAAAAT 2630
::: |||::|::| |||::|::|
478 Thr.....PheAspGlyAsn.....ThrAlaGlnThr 486
2631 CAGCGGCGGCAAGATACGGCATTTAAAGACAGCGAATGAGCG 2680
:::|||||::: |||::: |||:::
486 rAlaIaLYgLY.....AlaIleTyrrThrGlnThrGlnAspPheThrL 500
2681 TGCCGTCAAGCAGCAATTAGCAATTTAAACCTTGACAGCCCAACCTT 2730
|||::|::| |||::|::| |||::|::|
500 euThrGlySerThr.....GlyThrVal 507

2731 ACACITCAATTCGGCTATTCGCCAGATGGCGGCGCGCAACCGCGAC 2780
|||::|::| |||::|::| |||::|::|
508 ThrPheSerThr.....AsnThrAlaLysThrAlaGl 518
2781 TCGCACACATCGCGCGCGCGCGCGTTCGCGCGCGCTTCCTCAT 2830
::|::|::|::: |||::: |||:::
518 YAlaLeuHisThrLysGlyAsnThrSerPheThrLysAsnLysAlaLeuV 535
2831 TATCCGTTCACCGCCCAACTTCGTTGAATCCGTTTTCACACGCTGACG 2880
::: |||::|::| |||::|::| |||
535 al.....PheSerGlyAsnSerAlaThrAlaThrAlaThr 546
2881 GTAAACGCAAAATTTGAACGTCAGGACATTCGCTTATATGTGGAACT 2930
::: |||::: |||::: |||:::
547 ThrThrThrAspGlnGlnGlyCysGly.GlyAlaIleLeuCysAsnIleS 563
2931 CTTCGCTACCGCGCGACACAA...TTCAACTTGCGGAAAGTTCCGAA 2976
|||::|::| |||::|::| |||::|::|
563 eRguLserAspIleAlaThrLysSerLeuThrLysThrGlnAsnGluSer 579
2977 GGCACTTACACTTGGCGGTCAACAATACCGGCAACGAACTCGCAAC.. 3024
::: |||::: |||::: |||:::
580 LeuSerPhe.....IleAsnAsnThrAlaLysArgSerGlyGly 593
3025CTGAACAAATGAGGTAGTGGAGGAAGGAAGAACACAAACCGCTGT 3070
593 yIleTyrrAlaProLYgLYValIleSerGly.....S 604
3071 CCGAAACCTTAATTTACCTCGCAACAAAGCAACAGTCGATCGCGCGCG 3120
|||::|::| |||::|::| |||::|::|
604 eRguLserIleAsnPheAspGlyAsnThrAlaGluThrSerGlyGlyAla 620
3121 TGCGCTTACCACTCATCCGCAAAAGCGCGCACTTCGCTGATATATCC 3170
::: |||::: |||::: |||:::
621IleTyrrSerLysAsnLeuSerIleThrAlaAsnGlyPr 633
3171 GGTCAAGAACAAGAGCTTTCGCAACAATCGCGCAACGCAAGCAAGCAANA 3220
|||::: |||::: |||:::
633 oValSerPheThrAsnAsnSerGlyGlyGlyGlyAlaIleTyrrIleA 650
3221 AACAGCGGAAANAAACACACCGCAAAACCTTGACGCGCTGATTTGGC.. 3267
::: |||::: |||::: |||:::
3268 GCGGCGCGGATGCCGTGGAANAAGCAAGAAAGGTGGCGAACCAGCGCG 3317
::: |||::: |||::: |||:::
650 laSpSerLeYgluLeuSerLeuGluAlaIleAspGlyAspIleThrPhe 666
667 SerGlyAsnArgAlaThrGlnGlyThrSetThrProAsnSerIleHisLe 683
3318 GCAGCGAGCGGGGAAAATGTGGCATTTATGACGGCGGAGAAAGAAAA 3367
|||::|::| |||::|::| |||::|::|
683 uGlyAlaGlyAlaLysLleThrLysLeuAla.....AlaAla 695
3368 AACGGGTCAAGCGGATNAAAGACACCGCCCTTGGCGAAACGCGCAAGC 3417
|||::|::| |||::|::| |||::|::|
696 ProGlyHisThrIleTyrrPheTyrrAspProIleThrMetGluAlaProAl 712
3418 GAACCGCGCGGCTA..... 3433
::: |||::: |||::: |||:::
712 aSerGlylYthrIleGlnGluLeuValIleAsnProValValLysAlaI 729
3434CCACGCGCTTCCGCC..... 3448
|||::|::| |||::|::| |||::|::|
729 leValProProProGlnProLYgLYAsnGlyProIleAlaSerValProVal 745
3449 GCGCGCGCGCGCGCGCGGATTTGCGCACTCAACCCCAACCGCGAG 3498
::: |||::: |||::: |||:::
746 ValProValAlaProAla.....AsnProAsnThrGln 756
3499 CCCCAACCGCAGCGGACGACTGATCAAGCGCTTATGCAATAGCGGTTGAG 3548
::: |||::: |||::: |||:::
756 Y.....GlyThrVal 756

```

3549 TGAATTTCCGCCACGCTGTTTCCGGTACAGAGCAATAG 3598
757 .....ThrlleAlpheSer..... 761
3599 ACCCGATTATTCGCGAGACCGCGCAACCGCTTTGGACACAGCGCATC 3648
762 SerGlyLysLeuPro.....SerGlnAspAlaLeu 771
3649 CGGACACCAACACTACCTT.....CCGAGATTTCGGCGCCCA 3689
771 rleProAlaSerThrThrThrLeuAsnGlnLysIleAsnLeuAlaG 788
3690 CCGCCCAACACCGACCTGCGCAATCGGTATGCAAAAAACCTCGCA 3739
788 lylGlyAsn.....ValValLeuLys..... 794
3740 GCGGCGCGCTGCGATCCTGTTTCGCAACACCGGACCGCAACACCTTC 3789
795 GlnGlyAlaThrLeuGlnValTyrSerPheThr.....GlnGlnProAs 809
3790 GACGACGGCATCGGCACTGCGACGCGTTCGCCACGCGCGCTTTTCG 3839
809 pserThrVal.....PheMetAspAlaGlyThrThrL 820
3840 GCAATACGCGCATGACAGGTTCTACATGCGATACGCGCGCGCGGTT 3889
820 euglThrThrThrThrAsnThrThrAspGlySerIleAspLeuLysAsn 836
3890 TTAGACGCGCGGACGCTTCAGACGCGCATCGGCAAAATCGCGCGCG 3939
837 Leu.....SerValAsnLeu.....AspAlaLeuAspGlyLys.....Arg 848
3940 GTGCTCATTCACGCGCATTCAGCGCATACCGCGCGCTTCGCGGAT 3989
849 MetIleThrIleAlaValAsnSerThr.....SerGlyLysLe 861
3990 CGGACGCGAAGCGCATCGCGCAACGCGCTATTTCGTCMAAAAGCG 4039
861 ulpYsIleSer.....GlyAspLeuLysPheHisAsnGlnGlyLys 875
4040 ATTACGCGCTACGAAACGTCAATATGCGACCCCGCGCTTCATTCAC 4089
875 erPhe.....TyrAspAsn..... 879
4090 CGCTACGCGCGCGCATTTAGCATATTCATTCACACCGCGCAACA 4139
880 .....ProGlyLeuLysAlaAsnLeuAsnLeu..... 888
4140 CATTCATCAGCGCTTATTGAGCTGTCTATACGATGCGCGCTTCG 4189
889 .....ProPheLeuAsnLeuSerSerThr.....SerG 898
4190 GCAAAAGTCCGA.....ACACGCGCTCAATACCGCGCTATGCGT 4227
898 lYThrValAsnLeuAspAspPheAsnProIleProSerSerMetAlaAla 914
4228 CAGATTTTCGCG..... 4239
915 ProAspPyrGlyTyrGlnGlySerThrThrLeuValProLysValGlyAl 931
4240 .....AAACCGCGAGTGGGGAATGGCGGCTAAACGCGAATCA 4279
931 aglYglYlysValThrLeuValAlaGlnTyr.....GlnAlaL 944
4280 AAGGTTTCACG 4290
944 euglYTYrThr 947

```

seq_name: SwissProt_40:P2P_LACLC

seq_documentation_block: PRT; 1902 AA.

ID P2P_LACLC STANDARD; PRT; 1902 AA.

AC P15293;

```

DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE PII-type proteinase precursor (EC 3.4.21.96) (Lactocoeplin) (Cell wall-
DE associated serine proteinase) (LP151).
GN PRT.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid pUP763.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCDO 763;
RX MEDLINE=89313288; PubMed=2501630;
RA Kiwaki M., Ikemura H., Shimizu-Kadota M., Hirasima A.;
RT "Molecular characterization of a cell wall-associated proteinase gene
RT from Streptococcus lactis NCDO763."
RL Mol. Microbiol. 3:359-369(1989).
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some substrate preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC pro. in the P2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyse hemoglobin and oxidized
CC insulin B-chain.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-stb.ch).
CC
CC EMBL, X14130; CAA32350.1; -.
CC PIR, S06997; S06997.
CC DR HSSP: P00782; 2SHT.
CC DR MEROPS: S08.019; -.
CC DR InterPro: IPR001899; Gram_pos_anchor.
CC DR InterPro: IPR003137; PA.
CC DR Pfam: PF00746; Gram_pos_anchor; 1.
CC DR Pfam: PF02225; PA; 1.
CC DR Pfam: PF00082; Peptidase_S8; 3.
CC DR PRINTS: PR00723; SUBTILISIN.
CC DR PROSITE: PS00136; SUBTILASE_ASP; 1.
CC DR PROSITE: PS00137; SUBTILASE_HIS; 1.
CC DR PROSITE: PS00138; SUBTILASE_SER; 1.
CC DR PROSITE: PS00433; GRAM_POS_ANCHORING; 1.
CC DR PROSITE: PS00433; GRAM_POS_ANCHORING; 1.
CC DR Hydrolyase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
CC transmembrane.
CC KW SIGNAL.
CC FT 1 33
CC FT PROPEP 34 187
CC FT CHAIN 188 1902
CC FT DOMAIN 188 1876
CC FT TRANSMEM 1877 1895
CC FT DOMAIN 1896 1902
CC FT ACT_SITE 217 217
CC FT ACT_SITE 281 281
CC FT ACT_SITE 620 620
CC FT DOMAIN 1867 1872
CC FT SEQUENCE 1902 AA; 200139 MW; 4B8DBB64DB8CDF7 CRC64;

```

alignment_scores:

Quality: 199.50

Ratio: 0.360

Length: 1275

Gaps: 73

Percent Similarity: 43.451 Percent Identity: 21.098

alignment block:

US-09-303-518D-649 x P2P_LACLC ..

Align seg 1/1 to: P2P_LACLC from: 1 to: 1902

709 GATCATGGTGGTGCACACTTA..... 735
815 GlySerAspGlySerArgLeuAsnLeuProTyrMetGlyPhePheGlyAs 831
736 ...GGTGTGAAAAAATTAAACATPACCATATGAT.....TTTTCAC 775
831 PTPAsnAspGlyLysIleValAspSerLeuAsnGlyIleThrTyrSerP 848
776 CAACAGAGGCTCATTTGGCGACAGTGGCCACCAATGTTATCTATGAT 825
848 roAlaGlyLysAsnheGlyThr.....ValProLeu..... 858
826 GCCCAAAAGCAAAAGTGTAAATTAATGAGGATTCGCAACGCGCAACC 875
859LeuThrAsn.....LysAsnThrGlyThrG 867
876 CTATATAGCA.....AAAGCAATGGCTTCACAGCTGGTCGTA 913
867 nTyrTyrGlyGlyMetValThrAspAlaAspGlyAsnGlnThrVal... 882
914 AAGATTGGTTCTATGATGAATCTTTGCGAGATACCCATTCAGATTC 963
883AspAspGlnAlaIle... 887
964 TACGACCAACGTCACAAATGGGAAATACCTTTAACGACGATTAATATGG 1013
888AlaPheSerSerAspLys..... 893
1014 CACAGAAAAATCAATGCCAATGACACCAATCTCTCGGCTATATGAT 1063
894AsnAlaLeuTyrAsnAspI 900
1064 TAAAAACAGCAACCGTTCATATGTTTAATGTTTTCGAGACAGCA 1113
900 IeSerMetLysTyrTyrLeuLeuAsnIleSer..... 911
1114 AAGAAACCTGTTATCATGCTGACAGTGTGCAACAGTTATCAGCCAG 1163
911 911
1164 ACTGAATATGAGAAAAATATTCCTTT...ATTGACGAAGAAAAAGCG 1210
912AsnValGlnValAspIleLeuAspGlyGlnIle 923
1211 AATTGATACCTTACGACCAACATCAAGTGTGAGATTAATATTC 1260
923 snLysValThrThrLeuSerSerThrAsnAspGlyLysThrTyrTyr 939
1261CAAGGAGATTTTACGGTCTCGCGTGAATAATACGA 1295
940 AsnAlaHisSerGlnGlnTyrIleTyrGlnHisAlaPro..... 952
1296 AACTTGGCAAGC.....GGGCGGCTCATATCA 1324
953 .AlaTrpAspGlyThrTyrTyrAspGlnAspGlyAsnIleLysThrA 969
1325 GTGAAGACAGTACCGTTACTTGAAGATAACGCGTGGCAACGAC... 1371
969 IaaAspAspGlySerTyrThrTyrArgIleSerGlyValProGlnGly 985
1372CGCCTGTCCAAAATGGCGCA 1391
986 AspLysArgGlnValPheAspValProPheLysLeuAspSerLysAlaPr 1002
1392 AGGACCGCTCACGTT.....CAAGCCAAAGGGAAAAACCAAGGCTGCA 1435
...:|||||:|||||

1002 cThrValArgHisValAlaLeuSerAlaLysThrGluAsn..... 1015
1436 TCAGCGTGGCGACGCGTACAGTATTTGGATGACGAGCAGACAT... 1482
1016GlyLysThrGlnTyrTyrLeuThrAlaGlnAlaLysAspAsp 1029
1483AAGGCAAAAAACACCTTTTATGTAAT 1511
1030 LeuSerGlyLeuAspAlaThrLysSerValLysThrAlaIleAsnGluVa 1046
1512 CGCCTGGTCAAGCGGAGGGGTACGTCGAATGCAATGCGGATATCAG 1561
1046ThrAsnLeuAspAlaThrP 1053
1562 TCAACCCCGACAAACCTATTTGGCTTTGGCGGAGCGTTGGATTTA 1611
1053 heThr.....AspAla 1056
1612 AACGGGATTCGCTTCGTTCCACCGTATTCAAAT...ACCGATGA 1655
1057 GlyThrThrAlaAspGlyTyrThrLysIleGlnThrProLeuSerAspS 1073
1656 AGGCGCATGATGTTCAACACACATCAAGCAAAAGATCACCATTACCA 1705
1073 uGlnAlaGlnAlaLeuGly...AsnGlyAspAsnSerAlaGlnLeuTyrL 1089
1706 TTACAGGCATTAAGATATGCTACACCGGCAT... 1740
1089 eThrAspAsnAlaSerAsnAlaThrAspGlnAspAlaSerValGlnLys 1105
1741AACACAGCTTGGATGACAAAAAAGAAATGCGCTACACGTTG 1784
1106 ProGlySerThrSerPheAspLeu.....IleValAsnGlyGlyG 1119
1785 GTTGGCGGAAAGATACGACCAAAACGAGCGGCGTCAACCTGTTT 1834
1119 yLleProAspLysIleSerSerThrThrThrGly.....T 1131
1835 ACCAG..... 1839
1131 yGlnAlaAsnThrGlnGlyGlyThrTyrThrPheSerGlyThrTyr 1147
1840 CCCGCGCGAGAGACCGACCTGCTGCTCCGCGGACAAATTTAA 1889
1148 ProAlaIleValAspGlyThrTyrThrAspAlaGlnGlyLysHisAs 1164
1890 CGGCAACATCAG.....CAACAAACGCGCAACTGTTTTCAGC 1930
1164 PheAsnThrThrTyrAspAlaIleThrAsnSer.....PheThrAlaS 1179
1931 GCAGACCAACACCGCAC..... 1947
1179 erMetProValThrAsnAlaAspTyrAlaIleAlaGlnValAspLeuTyrAla 1195
1948GCCATCAATCATTTAAAGCACTTGGCGCAAAAGAGCGAT 1991
1196 AspLysAlaHisThrGlnLeuLeuLysHisPheAspThrLysValArgLe 1212
1992 T.....CCTCGCGGGAATTCGTGGGCAACAGACTGATCAACCGCA 2035
1212 uThrAlaProThrPheThrAspLeuLysPheAsnAsnGlySerAspGlnT 1229
2036 CATTTAAAGCGGAAACTTCGCAATTAAGCGGAGACAGCGGCGGTTTC 2085
1229 hrSerGluAla...ThrIleLysValThrGly.....ThrValSer 1241
2086 CGCAATGTTGCCAAAGTGA...GGCGATGCGCATTTGAGCAATCAGC 2132
1242 AlaAspThrLysThrValAsnValGlyAspThrValAlaIleLeuAspAl 1258
2133 CCAAGCAGTTTGTGTG.....GCACCGATCAAGACCCACACAA 2173
1258 aGlnHisHisPheSerValAspValProValAsnTyrGlyLysAsnThrI 1275

2174 TC...TGTACAGCTTCGACTGCAGCGGCTGTGACAAATTGTCTGAAAAA 2220
 ||| :||| :||| :|||
 1275 IelyValThrAlaThrAspIleuAsnThrThrThrGlutIleu 1291
 ||| :||| :||| :|||
 2221 ACCATACCAGC.....GATAAGTGATGGCTTCATTGACT... 2256
 ||||||| :||| :||| :|||
 1292 ThrIleThrSerSerTyraSPProAspMetLeuLysAsnSerValThrPhr 1308
 ||||||| :||| :||| :|||
 2257AAGCCGACATCACGCCGAATGTGATCTTCCGATCAGCTC 2299
 :||| :||| :||| :|||
 1308 eAspGIleuValThrPheGIleuAlaAsnGIleuPheAsnAlaThrSerAla 1325
 :||| :||| :||| :|||
 2300 ATTTA.....AATCTCACAGGCTTGCACACACTCAACGCGC..... 2334
 :||| :||| :||| :|||
 1325 ySPHeTyraSPProLysThrGIleuAlaThrIleThrGIleuLysValIys 1341
 :||| :||| :||| :|||
 2335AATCTTAGTGCAAATGCGCATACAGCTTAATCACTCAAG 2372
 :||| :||| :||| :|||
 1342 HisProThrThrThrLeuGIleuValAspIly...LysGIleuProIleu 1357
 :||| :||| :||| :|||
 2373 CCACAAGCCACCCAAAAGCGCACTAGCTC..... 2406
 :||| :||| :||| :|||
 1357 sAspSPleuThrPheSerPheThrLeuAspLeuGIleuThrLeuGIleu 1374
 :||| :||| :||| :|||
 2407GTGGCAATGCC.....CAAGCAACTTAAT 2433
 :||| :||| :||| :|||
 1374 ySPProPheGIleuValValGIleuAspThrThrGIleuAsnLysThrPheGIleu 1390
 :||| :||| :||| :|||
 2434 CAAGCC.....ACATTTAA 2447
 :||| :||| :||| :|||
 1391 GluAlaLeuThrPheIleLeuSPalaValAlaProThrLeuSerLeuAs 1407
 :||| :||| :||| :|||
 2448 CGGCAACACATCGGCT.....TCGGCAATGCTTCATTATTAATCA 2488
 :||| :||| :||| :|||
 1407 pSerSerThrAspAlaProValIyThrAsnAsnProAsnProGIleuThr 1424
 :||| :||| :||| :|||
 2489 GCGACCAAGCCGTAACAACAGCG.....AGTCTGACGCTTCCCGCG 2529
 :||| :||| :||| :|||
 1424 hngIyThrAlaThrAspAsnAlaGIleuThrLeuSerLeuSerIleuSmeIy 1440
 :||| :||| :||| :|||
 2530 AAGCTAGTGA....AACGTAGC..... 2550
 :||| :||| :||| :|||
 1441 SerSerValAlaSerGIleuIyThrValAspIleuAsnIleuAsnSerGIleuSPR 1457
 :||| :||| :||| :|||
 2551CATTCGCACTCAACGGTAATGTCTCCCTAGCCGATTAAGCGCAT 2596
 :||| :||| :||| :|||
 1457 oGIyHsmetAlaIleAspGIleuProValLysLeu..... 1468
 :||| :||| :||| :|||
 2597 TCCATTTTGAAGAAGCGCGCTTACCGCAATCAAGCGCGCAAGAT 2646
 :||| :||| :||| :|||
 1469LeuGIleuLysAsnVal 1474
 :||| :||| :||| :|||
 2647 ACAGCATTTACACTTAAAGACAGGAATGAGACGCTGCCGTCAAGCAGSA 2696
 :||| :||| :||| :|||
 1475 LeuThrValAlaValThrAspSerGIu..... 1483
 :||| :||| :||| :|||
 2697 ATTAGCAATTAAACCTTGACAAGCGCACCATTTACACTCAATTCGCGCT 2746
 :||| :||| :||| :|||
 1484AspAsnThrThrThrLysAsnIleThrValT 1494
 :||| :||| :||| :|||
 2747 ATCCGACGATGCGGAGGCGCAAACCGGACAGTGCAGATGCGCG 2796
 :||| :||| :||| :|||
 1494 yr.....TyrgluProLysThrLeuAlaIlaPro 1504
 :||| :||| :||| :|||
 2797 CGCGCGCTTCCGCGCTTCCGCGCTTCCCTATTATCGTTACACGCGC 2846
 :||| :||| :||| :|||
 1505ThrValThrProse 1509
 :||| :||| :||| :|||
 2847 AACCTGGTAGAATCCGTTTCAACAGCTGACGGTAACGCAAAATGTA 2896
 :||| :||| :||| :|||
 1509 rThrThrGIleuProAlaLysThrValThrLeuThrAlaAsnSerAlaIat 1526
 :||| :||| :||| :|||

```

1761 SerGlyGlnAlaSerAlaAspAlaSerAspYLeuAlaHisLeuGlnAl 1777
1778 GGTATGCAAAACCTGCGACGGCGCGCTGGCATCTGTTTTCGCA 3767
1777 AleuGlnSerLeuYstHrYstValAlaAlaValAlaGlnAlaAla 1794
3768 CAACCGACCGCAACACCTTCGACGACGATCGGCACTCGGCACGC 3817
1794 hrValGly...LysGlyAspGlyThrThrGlyThrSerAspLysGly 1809
3818 .....TTGCCACG..... 3826
1810 GlyGlnGlyThrProAlaProThrProGlyAspIleGlyLysAspLysG 1826
3827 .....GCCCGCTTTCGGCG.....AATACG 3848
1826 YAspGlyGlySerGlnProSerSerGlyGlyAlaProThrAsnProA 1843
3849 CATCGACAGGTTCTACATCG 3868
1843 LathrThrThrSerThrSer 1849

```

seq_name: Swissprot_40:N100_YEAST

seq_documentation_block:

```

ID N100_YEAST STANDARD: PRT: 959 AA.
AC 002629;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Nucleoporein NUP100/NSP100 (Nuclear pore protein NUP100/NSP100).
GN NUP100 OR NSP100 OR XKL068W OR XKL356.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
OC NCBI_TaxID:4932;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE-93054906; PubMed-1385442;
RA Wente S.R., Rout M.P., Blobel G.;
RT "A new family of yeast nuclear pore complex proteins.";
RL J. Cell Biol. 119:705-723(1992).
[2]
RN SEQUENCE FROM N.A.
RX STRAIN-S288C;
RX MEDLINE-94378724; PubMed-8091863;
RA Rasmussen S.W.;
RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
RT addition to seven ORFs with weak or no significant similarity to
RT known proteins.";
RL Yeast 10:S69-S74(1994).
-1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
-1- SUBCELLULAR LOCATION: Nuclear pore complex.
-1- DOMAIN: CONTAINS G-L-F-G REPEATS.
-1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: Z15035; CA78753.1;
DR EMBL: X75780; CA53406.1;
DR EMBL: Z28068; CA81905.1;
DR PIR: B44402; B44402.

```

```

DR PIR: S39173; S39173.
DR PIR: S44518; S44518.
DR SGD: S0001551; NUP100.
DR InterPro: IPR004325; Nucleoporein_FG.
DR Pfam: PF03093; Nucleoporein_FG; 24.
FT Nucleic protein; Transport; Repeat.
FT DOMAIN 33 571 G-L-F-G.
SQ SEQUENCE 959 AA; 99988 MW; D3985F9901BBA51 CRC64;

```

alignment_scores:

```

Quality: 195.50 Length: 1077
Ratio: 0.403 Gaps: 52
Percent Similarity: 45.032 Percent Identity: 19.591

```

alignment_block:

us-09-303-518d-649 x N100_YEAST

Align seq 1/1 to: N100_YEAST from: 1 to: 959

```

352 GCCTATACACAGTGTGATTTGGTGGGAA.....GGAAG 386
10 GlyGlySerAsnLeuSerPheGlySerAsnThrSerPheGlyGlyG 26
387 AATCCCGATCAACATCTTTACTTATTAATAATGTGAACGATTAAT 436
26 ngInSerGlnGlnProAsnSerLeuPheGlyAsnSerAsnAsnAsn 43
437 ATTAACGACGAGGACTAAGCCATCTCT...TATGGCGCGATATCAATG 483
43 snSerThrSerAsnAlaGlnSerGlyPheGlyGly..... 55
484 CCGCGTTTGCAATTAATTTGTCACAGATGACAGACCTGTGAATGACAG 533
56 .....PheThrSerAlaAlaGlySerAsnSerAsn 66
534 TTATATGATGATGGCGGCAATATATGATCAATAATTAATTAACCTGACCGTG 583
66 rLeuPheGlyAsnAsnAsn.....ThrGlnAsnAsn..... 76
584 TTCGATTGTTGGGACAGCAGCAATATTGGCATCTGATGAAGATGACCC 633
77 .....GlyAlaPheGlyInSerMetGlyAlaThrGlnAsnSerPro 90
634 AATAACCCGGAAGTTCTATATCATATTGCAAGTCGATTCTTGCTCGT 683
91 PheGlySerLeuAsnSerSerAsnAlaSerAsn..... 101
684 TGTGTCATATACCTTTGACAAATATGATCA.....GTGGTG 721
102 .....GlyAsnThrPheGlyGlySerSerMetGlySerPheGlyGly 117
722 GCACATGCACTTA.....GTAGTGAAAAAATTAACATACCCATAT 765
117 snThrAsnAlaAlaPheAsnAsnAsnSerAsnSerThrAsnSerProPhe 133
766 GGTATT..... 771
134 GlyPheAsnLysProAsnThrGlyGlyThrLeuPheGlySerGlnAsnAs 150
772 .....TTACCAAG 781
150 nasSerAlaGlyThrSerSerLeuPheGlyGlyGlnSerThrThr 167
782 GAGCGCATTTGGGACAGGCTCCACCAATGTTATTCATGATGCCAA 831
167 hrGlyThrPheGlyAsnThrGlySerSerPheGly..... 178
832 AACCAAAAGTGTTAATTAATGAGGATATTCAAACGACGACCCCTATAT 881
179 .....ThrGlyLeuAsnGlyAsnGlySerAsnIlePhe 189

```

```
882 AGNAAAAAGCAATGCTTCAGCTGCTGTAAGATGCTGTATGANG 931
    ||| :|||: |||
189 eglYAlaGlyAsnAsnSerGlnSerAsnThrThr.....GlyS 202
    :|||: |||: |||
932 AAATCTTGGTGGAGATACCATTCAGTATTCACGACCGCTCAAAAT 981
    :|||: |||: |||
202 erLeuPheGlyAsnGlnGlnSerSerAlaPheGlyThrAsnAsnGlnGln 218
    :|||: |||: |||
982 GGGAATATCTCTTTTAACGACGATATATATGACAGGAAAAATCAATGC 1031
    |||: |||: |||
219 GlySerLeu...PheGlyGlnGlnSerGlnAsnThrAsn.....AsnAl 232
    :|||: |||: |||
1032 CAACATGACACAAATCTCTGCTTAATACATTAATAAACGACACGCTTC 1081
    |||: |||: |||
232 aPheGlyAsnGlnAsnGlnLeuGlyGly..... 241
    :|||: |||: |||
1082 AATTTTAAATGTTTCTTATTCGACAGCAGACGACGCTTTATCAT 1131
    |||: |||: |||
242 .....SerSerPheGlySerLysProValGlySer 251
    :|||: |||: |||
1132 GCTGCA.....GCTGCTGTCACAGATTAATGACCCAGACTGAATATG 1175
    :|||: |||: |||
252 GlySerLeuPheGlyGlnSerAsnAsnThrLeuGlyAsnThrThrAsnAs 268
    :|||: |||: |||
1176 AGAAAAATATTTCTTTTATTGACGAAAGGCGAATTCATCTACTACCA 1225
    :|||: |||: |||
268 nATGAsnGlyLeuPhe.....GlyGlnMet.....A 277
    :|||: |||: |||
1226 GCAACATCAATCAAGGTGCT.....GGAGATTATATTTCCAGAGAT 1269
    :|||: |||: |||
277 snSerSerAsnGlnGlySerSerAsnSerGlyLeuPheGlyGlnAsnSer 293
    :|||: |||: |||
1270 TTACAGCTCGCT.....GAAATTAACGAACTTGCA 1304
    :|||: |||: |||
294 MetAsnSerSerThrGlnGlyValPheGlyGlnAsnAsnGlnMetG1 310
    :|||: |||: |||
1305 AGCGCGCGGCTTCATATCAGTGAAGACGATCCCTTACTTGAAAGTAA 1354
    |||
310 .....IleA 312
    :|||: |||: |||
1355 AGCGCGTGGCAACGACCGCTGTCGCAAAATCGCAGAA...GGCAGCGTG 1401
    |||: |||: |||
312 snGly.....AsnAsnAsnAsnSerLeuPheGlyLysAlaAsnThrPhe 326
    :|||: |||: |||
1402 CACGTTCAACCAAGGCGGAAACCAAGCTGATCAGCGTGGCGAGG 1451
    :|||: |||: |||
327 SerAsnSerAlaSerGlyGlyLeuPheGlyGlnAsnAsnGlnGlnGln 343
    :|||: |||: |||
1452 TACAGTATTTTGGATCAGCAGCAGCATTAAGCAAAAAACAAGCT 1501
    :|||: |||: |||
343 ySerGlyLeuPheGlyGlnAsnSerGlnThrSerGlySerSerGlyLeuP 360
    :|||: |||: |||
1502 TTACTGAATCGGCTTGTCAGCGGAGGCGTACGCTCACTGAATGCC 1551
    |||: |||: |||
360 heGlyGln..... 362
    :|||: |||: |||
1552 GATATATCAGTCAACCCCGACAACTATTTGGGCTTCGCGGCGAGG 1601
    :|||: |||: |||
363 AsnAsnGlnLysGlnProAsn..... 369
    :|||: |||: |||
1602 TTTGGATTAAAGCGGCTTCGCTTCACCGGATTCAAATACCG 1651
    :|||: |||: |||
370 .....ThrPheThrGln...SerAsnThrG 377
    :|||: |||: |||
1652 ATGAAGGGCGCATGATTTCTCAACCAATCAAGCAAAAGATCCACGCT 1701
    ||| :|||: |||
377 LylleGlyLeuPheGlyGlnAsnAsnAsnGlnGlnGlnSerThr... 392
    :|||: |||: |||
1702 ACCATTACAGCAATAAGATATTCCTACAAACGGCAATTAACAACACTT 1751
    :|||: |||: |||
393 GlyLeuPheGlyAlaLysProAlaGlyThrThrGlySer..... 405
    :|||: |||: |||
1752 GGATGCAAAAAAGAAATTCCTACAAACGTTGGTTGGCGAGAAAGATA 1801
    :|||: |||: |||

406 .....LeuPheGlyGlyAsnSers 412
    :|||: |||: |||
1802 CGAACAAAGACAGCGCGCTCAACCTGTTTACAGCCCGCCAGAA 1851
    :|||: |||: |||
412 erThrGlnProAsnSerLeuPheGlyThrThrAsnValProThrSerAsn 428
    :|||: |||: |||
1852 GACCGACCCCTGCTGCTTCCGCGACAAATTTAAACGGACATCAC 1901
    :|||: |||: |||
429 ThrGlnSer.....GlnGlnGlyAsnSerLeuPheGlyAlaThrLy 442
    :|||: |||: |||
1902 GCAACAAAGCGCAACACTGTTTTCAGCGCGACAGCAGACCGGCT 1951
    :|||: |||: |||
442 slLeuThrAsn.....MetProPheGlyGlyAsnProThrAlaAsnGlnS 457
    :|||: |||: |||
1952 ACAATCATTTTAAACGACCATTTGGTCGCAAAAAGAGGCAATTCCTCCGG 2001
    :|||: |||: |||
457 erGlySerGlyAsnSerLeuPheGlyThrLysProAlaSerThrThrGly 473
    :|||: |||: |||
2002 GAAATCGTGTGGACAAACGACTGATCAACCGCAATTTAAACGGGAAA 2051
    :|||: |||: |||
474 SerLeu...PheGlyAsnAsnThrAlaSerThrThrValProSerThrAs 489
    :|||: |||: |||
2052 CTTCCAATTTAAAGCGGACAGCGGCTGTTCCCGCAATGTTCCGCAAG 2101
    :|||: |||: |||
489 n...GlyLeuPheGlyAsnAsnAlaAsnAsnSerThrSerThrThrAsn 505
    :|||: |||: |||
2102 TGAAGCGGATTTGGCATTTGACATCAGCCGCAAGCAATTTTGGTGTG 2151
    :|||: |||: |||
505 hr.....GlyLeuPheGlyAla 510
    :|||: |||: |||
2152 GCACCGCATCAAGCCACACATCTGTACAGCTTCGAGACGAGGCTCT 2201
    :|||: |||: |||
511 LysProAspSerGlnSerLysProAlaLeuGlyGlyGlyLeuPheGlyAs 527
    :|||: |||: |||
2202 GACAAATTCGTGCAAAAACCATTAACGAGATTAAGTATGCTTCAT 2251
    :|||: |||: |||
527 nSerAsnSerAsnSerThrThrIleGlyGlnAsnLysProValPheGlyG 544
    :|||: |||: |||
2252 TGACATTAAGACCGACATCAGCGGCAATGTGATCTTCCGATCAGCTCAT 2301
    :|||: |||: |||
544 LylThrThrGlnAsn..... 548
    :|||: |||: |||
2302 TTAATCTCAAGGCTTCCACACTCAACGCAATTTAGTCAAAATGG 2351
    :|||: |||: |||
549 .....ThrGlyLeuPheGlyAlaThrGlyThr..... 557
    :|||: |||: |||
2352 CGATACAGCTTTATACAGTACGCCACAGCCGCAACCAAAAGGCACTTA 2401
    :|||: |||: |||
558 .....AsnSers 560
    :|||: |||: |||
2402 GCCTCGTGGCAATGCCCAAGCAACATTTAATCAAGCCACATTTAAAGGC 2451
    :|||: |||: |||
560 erAlaValGlySerThrGlyLysLeuPheGlyGln.....AsnAsn 573
    :|||: |||: |||
2452 AACACATCGGCTTCGGGCAATGCTTCATTTAATCTTAAGCCAGCCGCT 2501
    :|||: |||: |||
574 AsnThrLeuAsnValGlyThrGlnAsnValProProValAsnAsnThrTh 590
    :|||: |||: |||
2502 ACAAAAC.....GGCAGTGTGAGCTTTCGGCAACGCTAAGGCA 2542
    :|||: |||: |||
590 rGlnAsnAlaLeuLeuGlyThrThrAlaValProSerLeuGlnGlnAlaP 607
    :|||: |||: |||
2543 ACGTAAACCATTCGCACTCAGCAGGTAATGCTCCCTAGCCGATAGCA 2592
    :|||: |||: |||
607 roValThrAsnGlnGlnLeuPheSerLysLleSerLleProAsnSerLle 623
    :|||: |||: |||
2593 GTATTCATTTTGAAGCAGCGGCTTACCGGACAAATCAGCGCGGCAA 2642
    :|||: |||: |||
623 ..... 623
    :|||: |||: |||
2643 GGATACGGCATTAACCTTAAAGACAGCAATGAGCGCTCCGTCAGCA 2692
    :|||: |||: |||
```



```

624 .....ThrasProValLysa 629
2693 CGGAATTAGCAATTAAACCTTGAC.....AACGCCACCATTAACA 2733
      |||.....|||
629 lAtHrTherSerValasnaAlaasPmetLysArgasSerSerLeuThr 645
      |||.....|||
2734 CTCATTCGGCTTATCGCCACGATGCGGAGGCGCAACCGCGAGTGC 2783
      |||.....|||
646 .....SerAlaTyrArgLeu..... 650
2784 GACAGATCGCGCGCGCGCTTGCGCGCTTGCGCGCTTCCCTATTAT 2833
      |||.....|||
651 .....AlaProLys.....P 654
2834 CGGTACACCGCCCACTTCGTAGATCCGTTCAACAGCTGACGCTA 2883
      |||.....|||
654 rOLEuPheAlaProSerSer..... 660
2884 AACGCAATTTGAACGGTCAGGGAACATTCGCTTATGTCGCACTTT 2933
      |||.....|||
661 .....AsnGlyAspAlaLysPheGlnLysTyrPolLys..... 671
2934 CGGCTACCGGACGCAAAATTGAAGCTGGCGAAGTTCGGAAGGACCT 2983
      |||.....|||
672 .....ThleuGluArgSerAspArgLys 681
2984 ACACCTTGGGGTCAACATACCGGACGACGACCTGCAAGC..... 3024
      |||.....|||
681 eRTrHr.....SerAsnSerLeThrAspProLysSerTyrLeu 694
3025 .....CTGCAACAT 3035
695 AsnSerAsnAspLeuPheAspProAspArgTyrTyrLeuLysHisLe 711
3036 GAGGCTGTGGAAGGAAACACAAACCGCTGTCGAAACCTTAAT 3085
      |||.....|||
711 uValLys.....LysAsnAsnLys.....AsnLeuAsnArg 721
3086 TCACCTTGCMAAAGCAACACGTCGCGGCGGCGGCTTACCACTC 3135
      |||.....|||
721 alLys.....AsnHisAsnAspAspGlnAlaSerLysValLysLeu 734
3136 ATCCGCAAGACGCGGCTGCGCTCATTAATCCGTCACAAACAGA 3185
      |||.....|||
735 Val.....ThrPheThrThrGluSerAlaSerLysAspAspGln 747
3186 GCTTTCGACAACTCGGACAGGACGCAACCCAAACAGCGGCAAAAG 3235
      |||.....|||
747 nAlaSerSerSerLeuAla.....AlaSerLysLeuThrGlnLysA 761
3236 ACAACGCGCAAGCCTTGAACGCGCTGATGCGGCGGCGGATGCCGTC 3285
      |||.....|||
761 lAlHisSerProGlnThrAspLeu.....LysAsp..... 770
3286 GAAAGACAGAAAGCTGCGGACCGGCGCGGACGAGCGGAGAA 3335
      |||.....|||
771 AsPHisAspGlnSerThrProAspProLysSerSerProAsnGlySe 787
3336 TGTCGCATTTATGACGCGGAGGAGAAAGAA 3366
      |||.....|||
787 rThrSerLeuProMetLysGlnLys 797
seq_name: SwissProt_40: HXA3_HAEIN
seq_documentation_block:
ID HXA3_HAEIN STANDARD; PRT; 917 AA.
AC P45355;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme/hemopexin-binding protein precursor (Heme:hemopexin utilization
  protein A).
GN HXA.

```

```

OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NH1 N182;
RA MEDLINE=95115556; Pubmed=7815944;
RA Cope L.D., Thomas S.F., Latimer J.L., Slaughter C.A.,
RA Mueller-Eberhard U., Hansen E.U.;
RT "The 100 kDa heme:hemopexin-binding protein of Haemophilus
RU influenzae: structure and localization."
RU Mol. Microbiol. 13:863-873(1994).
CC -!- FUNCTION: BINDS HEME/HEMOPEXIN COMPLEXES.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 008349; AAA74139.1; -
KW Transport; Signal; Repeat.
FT SIGNAL 1 21
FT CHAIN 22 917 BY SIMILARITY.
FT DOMAIN 24 38 HEME/HEMOPEXIN-BINDING PROTEIN.
FT REPEAT 24 28 3 X 5 AA TANDEN REPEATS.
FT REPEAT 29 33 1-1.
FT REPEAT 34 38 1-2.
FT DOMAIN 111 668 6 X 6 AA APPROXIMATE REPEATS.
FT REPEAT 111 116 2-1.
FT REPEAT 103 208 2-2.
FT REPEAT 277 282 2-3.
FT REPEAT 399 404 2-4.
FT REPEAT 624 629 2-5.
FT REPEAT 663 668 2-6.
FT DOMAIN 159 170 2 X 6 AA APPROXIMATE TANDEN REPEATS.
FT REPEAT 159 164 3-1.
FT REPEAT 165 170 3-2.
SQ SEQUENCE 917 AA; 100148 MW; 7FF39BB8C046539D CRC64;

alignment_scores:
  Quality: 195.00 Length: 1095
  Ratio: 0.390 Gaps: 60
  Percent Similarity: 45.662 Percent Identity: 20.183

alignment_block:
US-09-303-518D-649 x HXA3_HAEIN
Align seg 1/1 to: HXA3_HAEIN from: 1 to: 917
442 GCAGGAGCTAAAGCCATCTTATGCGGC.....GATTATCATATGCC 485
      |||.....|||
21 AlaserThrGlnGlyLeuProGlnGlnGlyLeuProGlnGlnGlyLeuPr 37
486 GCGTTTGATTAATTTGTCACAGATGCAAGACCTGTTGAATGACGATT 535
      |||.....|||
37 GlnGlnLysPlyValValValGlyGlnAlaThrPheAspLysThr.... 52
536 ATATGATGCGGCGGAATATATGATGATGATTAATTAACCTGACCGTGT 585
      |||.....|||
53 .....lleAlaAspLysMetThrLleAsnGlnThrSer.....AspLysVal 66
586 CGTATTGGGCGGACGAGCAATATGCGCATCTGATGATGAGCCCA 635
      |||.....|||
67 GlnLleAsn.....TrpHisSer...PheAspLysGly 77
636 TAACCGGCAAGTTCAATATCATATTCGCAAGT.....GGGTATT 673
      |||.....|||

```

[illegible]

```
2096 CCAAGTGAAGCGCATTTGGCATTTGAGCAATTCAGGCCAGCAGCTTTT 2145
      :|||
      620 leasnlenuys..... 623
2146 GGTGTGCGACCGCATCAAGCCACACATCTGTACAGCTTGGAGTGGAC 2195
      :|||
      624 .....As 624
2196 GGGTCTGACAAATGTGTGCAAAAAACATTAACGAGATTAAGTGTG 2245
      :|||
      624 ngllyphevahlsleuenuahgluasnlllylsleuaspasn..... 638
2246 CTTCATTGACTAAGACGACATCAGCGGCAATGTGCAT..... 2283
      :|||
      639 .....SerlyValAspIleThrPheaspIysaspasnserserglnasp 652
2284 ...CTTGCCGATTCAGCTCATTTAAATTCACAGG...CTTCCCACT 2327
      :|||
      653 ThrleuAlaglnThrAsnAlrgleuGlyMetAsnGlylyValSerMetIl 669
2328 CAAGCGCAATCTTAGTCAAAATGCGCATACGCTTATACAGTCAGCACA 2377
      :|||
      669 eAsnSerHisIlelyslleValIglYaspIglYslglYlleserProt 686
2378 ACGGACCCCAAAAGCGCAACCTT.....AGCCTGCTGGCAATGCC 2418
      :|||
      686 hrglyThrlyTrAlatInMetPheleuIleGlyluLeuIleGlyluY 702
2419 CAAGCAACATTTAATCAAGC.....ACATTAACGG 2450
      :|||
      703 SerSerIlePheValIysSerHisIleGlylyTrIlePheIysTrhAsp 719
2451 CAACACATGGCGTTCGGC...AATGCTTCATTTATCTAAGCAGCAGCA 2497
      :|||
      719 yAspThrlyslleAlaGlyLysAsnSerlyslgluAspleuYslleThra 736
2498 CGGTACAAACGCGACACTGACGCTTCGCGCAACGCTTAAGCAACCTA 2547
      :|||
      736 laIleasnThrGly.....GlyArgAlaAlaGluGluAl 747
2548 AGGCATTCGCGACATCAGCGTATGTCTCCCTAGCCGATTAAGCAGTAT 2597
      :|||
      748 .....LeuIleAsnGlyAlaLeuGlySerAlaAspAsnAla.. 760
2598 CCATTTGAAAGCAGCGCTTACCGGCAATACGCGGCGGCAAGGATA 2647
      :|||
      761 .AsnIleAlaAsnMetAlaPheThrIleGlyAspSerAlaAsnThrlysr 777
2648 CGGCATTACACTTAAAGACAGCAGATGAGCGCTGCCGTGAGCAGCGAA 2697
      :|||
      777 hTrhIleGluAsnAlaAspIleThrAlaLeuAlaProasnGly..... 791
2698 TTAGGCAATTTAAACCTTGACACAGCCACCTTACACTC..... 2736
      :|||
      792 ...GlyThrAlaIlyrleuSerSerlyAspAlaIleGluVallyspr 807
2737 .AATTCGCGCATGCGCACGATGCGGAGGCGGCAAAACGCGAGTGGCA 2785
      :|||
      807 oAsnSerAsnPh.....Thrphep 814
2786 CAGATGGCGCGCGCGCTTCGCGCGCTTCGCGCGCTTCCTATATGCC 2835
      :|||
      814 hegluIleuProArg..... 818
2836 GTTACACCGCAACTTGGTAGATCCGTTTCAACAGCCTGACGGTAA 2885
      :|||
      819 .....GlyAsnLeuAsnGlnThrlyslleAs 828
2886 CGGC.....AAATGACCGTACAGGAGAACATTCGCTTATGTCGG 2926
      :|||
      828 nglYalaserThrlyleuSerleuIrglylPheAlaIrgleuYrAspL 845
2927 AACTTTGGCTACGCGACGCAAAATTAAGCTGCGGGAAGTTCGAA 2976
      :|||
```

```
845 yslleAsnGlyAlaIrgAlaserAsnleu..... 854
2977 GGCACCTACACCTTGGCGGTCAACATTAACGCGCAACGACCTGACGCC 3026
      :|||
      855 .....SerAl 856
3027 CGAACATTTAGCGTAGGAGGAGGAAAGCAACAAACCGCTTCCGAAA 3076
      :|||
      856 agluGlnleuAsnValThrAspAlaserGlu...LysIleIleAsnThrL 872
3077 ACCTTAATTCACCCCTGCAAAACGACACGTC..... 3108
      :|||
      872 yslleuValserSerleuAspValGluYslleuValserValAlaValCys 888
3109 GATCGCGCGCGGTGGCTTACCACTCATCTCCGCAAGACGCGAGTTCGG 3158
      :|||
      889 AspAlaGly..... 891
3159 CTTCGATTAATCCGCTGCAAAAGCAAGAGCTTCCGACAAA..... 3198
      :|||
      892 .....AsnGlyCysgluGlnGlnIlePheGlyAspIysGlyAsnAsnT 906
906 hrlYsValserValIglYgluLeuGluAlaGluGln 917
seq_name: swissProt_40:PM20_CHLPN
seq_documentation_block:
ID PM20_CHLPN STANDARD: PRT; 1723 AA.
AC Q92812; Q9K2C1; Q9RB59;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Probable outer membrane protein pmp20 precursor (Polymorphic membrane
protein 20).
OS PMP20 OR CPN0540 OR CP0212.
GN Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
Olinger U., Grimwood J., Davis R.W., Stephens R.S.;
"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
White O., Hickey B.K., Peterson J., Utterback T., Berry K., Bass S.,
Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
Gwin M., Nelson W., Deboy R., Kolony J., McClarty G., Salzberg S.L.,
Eisen J., Fraser C.M.;
"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39.";
Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
Shida T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
"Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA.";
Nucleic Acids Res. 28:2311-2314(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
```


[illegible]

2314 GGGCTTGGCACACTACAGCGCAATCTTAGTGCAATGGCGATACACGTTA 2365
 1050 AspaSnSer:::|||||::: ||::: ||::: ||::: ||
 2364 TACAGTCAAGCCACAGCGCACCCCAAAAGCGCAC..... 2397
 1066 pLseSerTrgThrGlyAlaThrPheIleGlyAsnSerSerLysHisAspG 1083
 2398CTTACCCTCGTGGGCAAGGCCCAA 2421
 1083 LysErAlaIleCysCysSerThrAlaLeuThrIleAlaIleProHisSerGln 1099
 2422 GCACAACTT.....AATCAAGCCACTTAAAGCGCAACACATCGGCTTC 2465
 1100 LeuIlePheGlnAsnHisValThr.....GluThrThrAlaThr 1113
 2466 GCGCAATGCTTCACTTAAAT..... 2484
 1113 rThrLysAlaSerIleLysAsnHisLeuAlaIleTyrGlyAsnAsnG 1130
 2485 ..CTAACGCAACAGCCGCTA.....CAAAAGCGCAGTCTGACG 2520
 1130 LuthrSerAspIleThrIleSerLeuSerAlaGlnAsnGlySerIlePhe 1146
 2521 CTTTCCGGGCAC..GCTAAGCGCAAGCGTAAAGCAATTCGCACTCAACGG 2567
 1147 PheLysAsnHisLeuSerThrAlaThrAsnLysTyrCysSerIleAlaI 1163
 2568 TAAATGCTCCCTA.....GCCATTAAGCGAGTATTCG 2599
 1163 yAsnValLysPheThrAlaIleGlnAlaSerAlaGlyLysAlaIleSerP 1180
 2600 AATTGTAAGCAAGCCGCTTACCGCAATACAGGCGCGCAAGCATACG 2649
 1180 heTyrAspAlaVal.....AsnValSerThrLysLysLuthrAsn 1192
 2650 GCATTTACACTTAAAGACAGCGCAATGACGCTGCCGTACAGCGCAAGTAT 2699
 1193 AlaGlnGlnLeuLysLeuAsnGlnLysAlaThrSerThnGlyThrIleLe 1209
 2700 A.....GGCAATTTAAACCTTGACACAGGCCACCTTACACTCAATTCG 2743
 1209 uPheSerGlyGlnLeuHisGlnAsnLysSerTyrIleProGlnLysValT 1226
 2744 CCTATCGGCAC.....GATCGCGCAGGCGGCCCAACG 2775
 1226 hrPheAlaHisGlyAsnLeuIleLeuGlyLysAsnAlaGlnLeuSerVal 1242
 2776 GCGAGTGGCAGACAGTGGCGCGCGCGCGTGGCGCGCTTCGCCCGCTTC 2825
 1243 ValSerPheThrGlnSerProGlyThrThrIleThrMetGlyProGlySe 1259
 2826 CCTATTATCC..... 2835
 1259 rValLeuSerAsnHisSerLysGlnAlaGlyGlyIleAlaIleAsnAsnV 1276
 2836GTT 2838
 1276 alIleIleAspPheSerGlnIleValIleProThrLysAspAsnAlaThrVal 1292
 2839 ACACCGCGCACT..TCGGTAGAATCCCGTTTCAACACAGCTGACGGGTAA 2885
 1293 AlaProThrThrLeuLysLeuValSerAlaThrAsn..AlaAspSerLys 1308
 2886 CGGCAAAATTGAAGCGTCAGGGAACATTCGCGTTTATGTGGAACCTCTTCG 2935
 1309 AspLysIleAspIleThrGlyThrValThrLeu..... 1319
 2936 GCTACCGCAGCGACAAATTTGAAGCTGGCGGAAAGTTCGCAAGCAGCACTTA 2985
 1320LeuAspProAsnLysAsnLeuTyrGlnAsnSerT 1331
 2986 ACCTTGGCGGTCAACAATACGGCAACGACACTGCAAGCCTTCGAACAACTT 3035

[illegible]

```

567 spherSerGluAspIleGlySerGlnSerArgIleAsnThrValArgLeuG 584
      :::::::::: :::::::::: :::::::::: ::::
1870 .....TCGGCGGAGCAAAATTAAAGCGC... 1893
584 IuThrGlyThrArgSerIleTyrSerGlyGlyValLysPheLysGlyGly 600
1893 ..... 1893
601 GluLysLeuValIleAsnSpherIleTyrAlaProTyrAsnTyrPheAs 617
1894 .....ACATCAGCGCAACAAAGCGCAACTGTTTCA 1927
617 pAlaArgAsnIleLysAsnValGluIleThrAsn...LysLeuAlaPheG 633
1928 GCGGAGACCAACACCGCAGCGC.....TACATCATTTA 1962
      :::::::::: :::::::::: :::::::::: ::::
633 LyrGlnGlySerProTyrGlyThrAlaLysLeuMetPheAsnSnuLeu 649
1963 AAGGACCATTTGGTCCAAAAGAGGCGATTCTCGGGGAATCGTGTG 2012
      :::::::::: :::::::::: :::::::::: ::::
650 Thr.....LeuGlyAsnAlaValMetAs 658
2013 GGACAAAGCTGGATCAACCGCACATTTAAAGCGAA..... 2049
      :::::::::: :::::::::: :::::::::: ::::
658 pTyrSerGlnPheSerAsnLeuThrIleGlnGlyAspPheValAsnAsnG 675
2050 .....ACCTCCAAATTTAAAGCGGAGCGCGGTGTTCCGC 2088
      :::::::::: :::::::::: :::::::::: ::::
675 lngIYThrIleAsnTyrLeuValArgIleGlyGln...ValAlaThrLeu 690
2089 AATGTTGCCAAAGTGAAGAGCGATTGGCATTTGACCAATCAAGCCCAAGC 2138
      :::::::::: :::::::::: :::::::::: ::::
691 AsnValGlyAsnAlaIleAlaIleMetPhePheSerAsnValAspSerAl 707
2139 AGTTTGGTGTGCACCGCGATCAAGCCACACAAATCTGTACAGTTCCG 2188
      :::::::::: :::::::::: :::::::::: ::::
707 aThrGlyPheTyrGlnProLeuMetLys..... 716
2189 ACTGAGCGGCTGTGCAAAATTTGTTCGAAAAACCATTTACGACGATTA 2238
      :::::::::: :::::::::: :::::::::: ::::
717 .....IleAsnSerAlaGlnAspLeuIleLysAsnLysGlu 728
2239 GTGATGCTTCATTGACTGAACCGACACATCAGCGGCAATGCGATTGCG 2288
      :::::::::: :::::::::: :::::::::: ::::
729 HisValLeuLeuLysAlaLysIleIleIleTyrIleLysValSerIleAl 745
2289 CGATCAC.....GCTCATTTAAATCTC.....ACAGGC 2317
      :::::::::: :::::::::: :::::::::: ::::
745 yThrAspSerIleAlaHisValAsnLeuIleGluGlnPheLysGluArgL 762
2318 TTGCCACACTCAACGCGCAAT.....CTTAGTGCAT 2349
      :::::::::: :::::::::: :::::::::: ::::
762 euAlaLeuTyrAsnAsnAsnAsnArgMetAspIleLysValAlaArgAsn 778
2350 GCGGATACACGTTATACAGTACGCCACACGCCACCAAAAGCGCAACT 2399
      :::::::::: :::::::::: :::::::::: ::::
779 ThrAsp.....AspIleLysAlaCysGlyThrAlaIleGlyAsnGl 792
2400 TACCGCTCGCGCAATGCCCAACACATTTATTCAGGACCATTAAGC 2449
      :::::::::: :::::::::: :::::::::: ::::
792 nSerMetValAsnAsnProGlu.....AsnTyrLysTyrLeuGluG 806
2450 GCAACACA.....TCGGCTTCGGGCAATGCTTCATT 2481
      :::::::::: :::::::::: :::::::::: ::::
806 LysAlaIleThrLysAsnIleGlyTyrLeuTyrThrAlaAsnGlySer... 821
2482 AATCTACAGCAACCGCGGTACAAAGCGCACTGACAGCTTCGCGCA 2531
      :::::::::: :::::::::: :::::::::: ::::
822 LysIleSerValHisTyrLeuGlyAsnSerThrProThrGluAsnGlyG 838
2532 CGCTAAGCAAAAGCTAAGCCATTCGCACTCAACGTTATGTCCTCGT 2581
      :::::::::: :::::::::: :::::::::: ::::

```

```

838 yAsnThrThrAsnLeuProThrAsnThrAsn...LysValArgPhe 854
2582 CGGATAGGAGATATTCATTGGAAGCAGCCGCTTACCGACAATC 2631
854 lAsnThrAlaLeu...lIeLysAsnAlaProPheAlaArgTyrSer 868
2632 AGCGGGGCAAGGATGCGCATTTACACTTAAAGACAGGATGAGCGT 2681
869 AlaThrProAsnLeuValAlaIleAsnGlnHisAspPhe...GlyThrI 884
2682 GCGGTGCGACGAGATTAAGCAATTTAACTTGACAAAGCCACCATTA 2731
884 eGluSerValPheGluLeuAlaIleAsnArgSerAsnAspIleAspThrLeuT 901
2732 CACTCAATTCGCGCTATCGC..... 2751
901 yAlaAsnSerGlyAlaGlnGlyArgAspLeuGlnThrLeuLeuIle 917
2752 .....CAGCATGCGGCGAGGGCGCAACCGGACGATGCGACAGATCGCC 2795
918 AspSerHisAspAlaGlyTyrAla..... 925
2796 GCGCGCGCGTTCGCGCGCTTCGCGCGCTTCCTATTAATCCGTTACACCGC 2845
926 .....ArgThrMetIleAspAlaThrSera 934
2846 CACTTCGCTGAGATCCCGTTTCAACACGCTGACGTAACGCGCAATTCG 2895
934 lAsnGlnIleThrLysGlnLeuAsnThrAlaThrThr...Leu 948
2896 AACGCTCAGGCAACATTCGCTTATGTGGAAGCTTCGCGTACCGCAG 2945
949 AsnAsnIleAlaSer..... 953
2946 CGACAATTTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACTTGGCGG 2995
954 .....LeuGlnHisLysThrSerGlyLeuGlnThrLeuSerL 966
2996 TCACAATACCGGCAACGAACTGCAAGCTCGAACAATGACGTTAGTG 3045
966 euserSn..... 968
3046 GAAGAAAACACAAACCGCTGCTCCGAAACCTTAATTTACCCCTGCA 3095
969 .....AlaMetIleLeuAsnSerArgLeuVal 977
3096 AAC.....GACACGCTGCATGCCGCGGCTGCGCTT 3127
977 lAsnLeuSerArgArgHisThrAsnHisIleAspSerPheAlaLysArgL 994
3128 ACCAATCTATCCGCAAAAGCGGAGTTCGCGCTGCATATCCGCTCANA 3177
994 euglnAlaLeu.....Lys 998
3178 GAACAAGAGCTTTCGACAAACTCGGCAAGCAGAA..... 3213
999 AspGlnArgPheAlaSerLeuGlnSerAlaIleGluValLeuTyrGlnP 1015
3214 .GCCAAAACAGCGCGGAAAAAGACACGCGGCAAGCTTGAGCGCTGA 3262
1015 eAlaProLysTyrGlnLysProThrAsnValTyrPala..... 1027
3263 TTGCGCGCGGCGGCGATCGCTCGAAAAAGACAGAAAGCGTTCGCGCAACG 3312
1028 .....AsnAlaIleGlyGlyThr 1033
3313 GCCCGGACGAGCGCGGGAATGTGGCATTTGACAGGCGGAGAGA 3362
1034 SerLeuAsnSerGlyGly...AsnAlaSerLeuTyr..... 1044
3363 GAAAAACGCGTGCAGCGGATTAAGACACGCGCTTGCGGAAACAGCGCG 3412
1045 .....GlyThrSerAlaGlyValAlaSpAlaTyrLeuAsnGlnGlyValG 1059

```

```

3413 AACGGAAACCCGGCGGCTACACACCGGCTTCCCCGGCGCGCGCGCC 3462
1059 lAlaL..... 1060
3463 CGCGGGGATTTGCGCAACTGCACCCACCGCAGCGCCCAACCGCAGG 3512
1060 ..... 1060
3513 CGACGTATCAGCCGTTATGCCAATAGCGTTTGTAGTAATTTCCGCA 3562
1061 .....lIeValGlyGlyPheGlySerTyrGlyTyrSerPheSerAsnG 1076
3563 CGCTCAACAGC..... 3573
1076 lAlaAsnSerLeuAsnSerGlyAlaAsnAsnThrAsnPheGlyValTyr 1092
3574 .....GTTTGGCGCTACAGCAGCAATTAAGC.....CGCGT 3605
1093 SerArgIlePheAlaAsnGlnHisGluPheAspPheGluAlaGlnGlyAl 1109
3606 ATTGGCCGAGACCGCGCAACGCGCTTGGACACAGCGC...ATCCGGG 3652
1109 aLeuGlySerAspGlnSerSerLeuAsnPheLysSerAlaLeuAlaArg 1126
3653 AC.....ACCAACAGC 3663
1126 sPheAsnGlnSerTyrAsnTyrLeuAlaTyrSerAlaAlaThrArgAla 1142
3664 TACCGTTGGCAAGATTCGCGCGCTACCGCAACAAACGACCTGGCGCA 3713
1143 SerTyrGlyTyrAspPheAlaPhePheArgAlaLeuValLeuLysP 1159
3714 ATTCGATGACGAAACCTCGGCGAGCGCGCGCTGCGCATCTGTTT 3763
1159 O.....SerValGlyValSerTyr 1166
3764 CGCACACCGGACCGAAACACCTTCGACGAGCATCGCACTCGCA 3813
1166 snHisLeuGlySerThrAsn...PheLysSerAsnSerAlaGlnLysVal 1181
3814 CGGCTGCGCCAGCGCGCGTTTTCGGGCAATACGACATCGACAGTTCTA 3863
1182 AlaLeuLysAsnGlyAla..... 1187
3864 CATCGCATCAGCGCGCGGCGGTTTACGACGCGACCTTCAGAGC 3913
1188 .....SerSerGlnHisLeuPheAsnA 1195
3914 GCATCGAGAGCAAAATCCGCGCGCGCTGCTGCATTTACGGCATTCAGGCA 3963
1195 lAsnAla.....AsnValGlnAla 1201
3964 CGATACCGCGCGGTTTCGGGCGATTCGGCATCGAACCGCATCGCGCGC 4013
1202 ArgTyr.....TyrTyrGlyAs 1207
4014 AACCGCTATTTGCTCCAAAAGCGGATTAACGCTTCGAAACGTCATA 4063
1207 pThrSerTyrPheTyrMetAsnAlaGlyValLeuGlnGluPheAlaAsn 1224
4064 TC...GCCACCCCGCGCTTCGATCAACCGCTTACCGC 4098
1224 heGlySerSerAsnAlaValSerLeuAsnThrPheLys 1236

```

seq_name: SwissProt_40:PM2L_CHLPPN

seq_documentation_block:
 ID PM2L_CHLPPN STANDARD; PRT; 1609 AA.
 AC Q926D5; Q9RB58;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Probable outer membrane protein pmp21 precursor (polymorphic membrane
protein 21).
GN PMP21 OR CFN0963 OR CP0897.
OC Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. (1)
OX NCBI_TaxID=83558;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-CWL029.
RX MEDLINE=99206606; PubMed=10193388;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.",
RT Nat. Genet. 21:385-389(1999). (2)
RN
RP SEQUENCE FROM N.A.
RC STRAIN-AR39.
RX MEDLINE=20150255; PubMed=10644935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uppertack T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dolson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39.",
RT Nucleic Acids Res. 28:1397-1406(2000). [3]
RN
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kunara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.",
RL Nucleic Acids Res. 28:2311-2314(2000).
CC
CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC
CC -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC
CC
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001676; AAD19099.1; -
CC EMBL; AE002248; AAF38684.1; -
CC EMBL; AP002548; BAA99171.1; -
CC PHC1-2DPAGE; Q9RB58; -
CC TIGR; CP0897; -
CC InterPro; IPR003357; OMP.
CC Pfam; PF02385; OMP; 1.
DR
DR Outer membrane; Signal; MultiGene family; Complete proteome.
KW SIGNAL 1 30
FT CHAIN 31 1609 POTENTIAL.
FT CONFLICT 420 420 I -> M (in ref. 3).
FT SEQUENCE 1609 AA; 170865 MW; 2604C3E97C4024CB CRC64;

| | | |
|---------------------|--------|--------------------------|
| alignment_scores: | | |
| Quality: | 194.50 | Length: 14111 |
| Ratio: | 0.332 | Gaps: 76 |
| Percent Similarity: | 41.531 | Percent Identity: 20.051 |

alignment_block:
US-09-303-518D-649 x PM21_CHLPN .,

Align seg 1/1 to: PM21_CHLPN from: 1 to: 160e

1108 ACAGCAAGA.....GAACCTGTTATCATGCTGCAGGTGCTCAACAG 1151
||||||| . ||||| :|||

[illegible]

2556 CGCACACCAACGGTAATGTCCTCCCTACCGCATATAGCAGATTCACATTTGG 2605
 838GNGlyAsnIleAlaPheLysGluAsnPheValPheGlySerC 852
 2606 AAAGCAACGCC..... 2616
 852 LuAsnGlnArgSerGlyGlyValIleLeuAlaIleLeuAlaAsnSerSerValAsn 868
 2616 2616
 869 IleGlnAspAsnAlaGlyAspIleLeuPheValSerAsnSerThrGlySe 885
 2617TTTACCGGACAA..... 2628
 885 rTyrglyValAlaIlePhePheValGlySerLeuValAlaSerGluGlySerA 902
 2629ATCAGCGCGGCAACGATACGCGCATTAACCTTA 2661
 902 snProArgThrLeuThrIleThrGlyAsnSerGlyAspIleLeuPheAla 918
 2662 AAAGCAACCGCAATGAGC..... 2679
 919 LysAsnSerThrGlnThrAlaAlaSerLeuSerGluLysAspSerPheG 935
 2679 2679
 935 yGlyValAlaIleTyThrGlnAsnLeuLysIleValLysAsnAlaGlyA 952
 2680CTGGCGGTACGACGCGAATTA..... 2700
 952 snValSerPheTyrglyAsnArgAlaProSerGlyAlaGlyValGlnIle 968
 2700 2700
 969 AlaAspGlyGlyThrValCysLeuGluAlaPheGlyGlyAspIleLeuPh 985
 2701GGCAATTAAACCTTGAC..... 2718
 985 eglGlyAsnIleAsnPheAspGlySerPheAsnAlaIleHisLeuGlySG 1002
 2719AACCCACCACTTACACTCAATTCGCGCTATCGC..... 2751
 1002 LysAsnSperLysIleValGluLeuSerAlaValGlnAspLysAsnIle 1018
 2752CACGATCGGCGACGGGCGGCAACCGGACGACAGATGCGCC 2795
 1019 IlePheGlnAspAlaIleThrTyrgluGlnAsnThrIleArgGlyLeuPr 1035
 2796 GCGCGCGCGCTTCGCGCGCTTCGCGCGCTTCCTATTATCCGTTACCGC 2845
 1035 OAspLys.....AspValSerProI 1042
 2846 CAACCTTCGGTAGAATCCCGTTTCAACACGCGTGAACGGTAACGCCAAATTG 2895
 1042 euSerAlaProSerLeuIlePheAsnSerLysProGlnAspAspSerAla 1058
 2896 ...AACGGTCGGAGCAATTCGCGTTATGTGGAACTCTTGCGCTACCG 2942
 1059 GlnHisHisGluGlyThrIleArgPheSerArgGlyAla..... 1071
 2943 CAGCGACAAATTTGAACTGCGCGGAAATTCGGAAGGCACTTACACTTGG 2992
 1072 .SerLysIleProGlnIleAlaIleAlaIleGlnGluGlyThrLeuAlaLeuS 1088
 2993 CGGTCAAC.....AATACGCGCAAC 3012
 1088 eArgLysAlaGluLeuThrPheLysAlaGlyLeuLysGlnGluThrGlySer 1104
 3013GAACCTGCAAGCCTCGACAATTTGACGATGTGGAAG 3050
 1105 SerIleValLeuSerAlaGlySerIleLeuArgIlePheAspSerGlnVa 1121
 3051 AAAGCAACAACACCGGTG...TCGGAACACTTAATTCACCTGCAAA 3097

```

1121 1aspserSerAlaProLeuProThGluAsnLysGluGluThreU... 1136
1308 AGAAGACAGTTCGATGCGGCGCGGCGGCTTACCACTCATCCGCAAGAC 3147
1137 .....ValSerAlaGlyVal..... 1141
3148 GCGGAGTTCGCGCTGCATATCCGGTCAAGAACAGAG..... 3186
1142 ..GlnIleAsnMetSerProThrProAsnLysAspLysAlaValAs 1157
3187 .....CTTCGCAACAACTCGCGCAAGCGCAAA.....GCCAANA 3220
1157 pThProValLeuAlaAspIleIleSerIleThValAspLeuSerSer 1174
3221 AACAGGCGGAAAAAGACAGCGCAAGCGCT.....GACGCGCGGAT 3264
1174 heValProGluGlnAspGlyThrLeuProLeuProProGluIleIle 1190
3265 GCGGCGCGGCGCGCGATGCGCTGAAAGACAGAAAGCGTTGCGCAACGCG 3314
1191 ProLysGly.....ThrLysLeuHisSerAsnAlaIleAspLe 1203
3315 CCGGACGAGCAGCGGCGGAAATGTCGGCATTTAGCAGCGCGAGAAAGAG 3364
1203 uLysIleIleAspProThrAsnValGlyTyr.....GluAsnH 1216
3365 AAAAAGCGGTGACAGCGCGATAAAGACACCGCTTGGCAACAGCGCGGA 3414
1216 IsAlaLeuLeuSerSerHisLysAspIleProLeuIleSerLeuLysThr 1232
3415 GCGGAAACCGCGCGGCTAC..ACCGCTTCCCGCGCGCGCGCGCGC 3461
1233 AlGlnGlyMetThrGlyThrProThrAlaAspAlaSerLeuSerAsn11 1249
3462 CCGCGCGGAT.....TTCGCGCAACTGCACCC..... 3489
1249 eLysIleAspValSerLeuProSerIleThrProAlaThrGlyHisT 1266
3489 ..... 3489
1266 hGlyValTrpSerGluSerLysMetGluAspGlyArgLeuValGly 1282
3490 ...CAACCG.....CAGCGCAACCGCGAGCGCGCGCTGAT 3521
1283 TrpGlnProThrGlyTyrLysLeuAsnProGluLysGlnGlyAlaLeuVa 1299
3522 CAGCGCTTATGCCAATAGCGGTTTGAATTTCCGCGCACGCTCAACA 3571
1299 LeuAsnAsnLeuTrpSerHisTyrThAspLeuAlaGlyAlaLeuLysGln 1316
3572 GCGTTTGGCGGTACAGAGCAATTAAGCGCGCTATTTGCCGGAAGACGCG 3621
1316 IuIlePheAlaHisHisThrIleAlaGlnArgMetGluLeuAspPheSer 1332
3622 CGCAACGCGGTTTGGACAGCGGATCGCGGACCAACACATACGCTTC 3671
1333 ThrAsn...ValTrpGlySerGlyLeuGlyValValGluAspCysGlnAs 1348
3672 G...CAAGATTTCCGCGCTACCGCGCAACAACCGAGCTCGCGCAATCG 3718
1348 nIleGlyLeuPheAspGlyPheLysHisHisLeuThGlyTyrAlaLeuG 1365
3719 GTATGCAAAAAACCTCGCAGCGG...CGCGTGGCATCCGTTTGG 3765
1365 LysLeuAspTrhGlnLeuValGlnAspPheLeuIleGlyLysPheSer 1381
3766 CAC.....AACGAGACGGAACACCTTCGACGAGCGGATCGCAACTC 3809
1382 GlnPheGlyLysThrGlnSerGln...SerTyrLysAlaLysAsnAs 1397
3810 GGCACGCGCTTCCACGCGCGCTTTTCGGCAATACGCGCATCGACAGST 3859

```

```

1397 pValLysSerTyrMetGlyAlaAla..... 1405
3860 TCTACATCGGATCAGCGCGCGCGGCTTTTACAGCGCGACGCTTCA 3909
1406 ..TyrAlaGlyIleLeuAlaGlyProTrpThrLeuLysGlyAlaPheVal 1421
3910 GACGCGATCGAGGCAAAATCCGCGCGCGCTGCTGATTAAGCGCATTCG 3959
1422 TyrGly.....AsnIleAs 1426
3960 GGCACGATACCGCGCGCTTTCGCGGATTCGCGCATCGAACCGACATCG 4009
1426 nAsnAspLeuThrThrAspTyrGlyThrLeuGlyIleSerThGlySerT 1443
4010 GCGCAACGCGCTATTTCGTC.....CAAAAGCGGATTAACGCTACGAA 4053
1443 rPileGlyLysGlyPheIleAlaGlyThrSerIleAspTyrArgTyr... 1458
4054 AACGTCATATTTCGCGCACCGCGCGCTTCGATTCACCGCTACCGCGCGG 4103
1459 .....IleValAsnPro.....ArgArgPh 1465
4104 CATTAAGCAGATTTTCAATTCACACCGCGCACACATTTTCATCAGCGC 4153
1465 eIleSerAlaIleValSer.....ThrValValP 1475
4154 CTATTATTAGCGCTGCTTCATACGAT.....GCCGT 4185
1475 roPheValGluAlaGluTyrValArgIleAspLeuProGluIleSerGlu 1491
4186 TCGGCGCAAA...GTCCGACACCGCTCAATACCGCGCTATTGGCTCAGA 4232
1492 GlnGlyLysGlyValAlaArgThr..... 1498
4233 TTTCGCAAAACCGCGATGCGCAATGGCGGTAAAGCGCAATTCANAAG 4282
1499 .PheGlnLysThrArgPheGluAsnValAlaIle.....P 1510
4283 GTTTCACGCTGTCCTCCACGCTGCGCGCGCAAGCGCGCACTGGAA 4332
1510 roPheGlyPheAlaLeuGlnHisAlaTyrSerArgGlySerArgAlaGlu 1526
4333 GCGCAACACAGCGCGGATCAATTAAGCTAC 4365
1527 ValAsn.....SerValGlnLeuAlaTyr 1534

seq_name: SwissProt_40:P2P_LACPA

seq_documentation_block:
ID P2P_LACPA STANDARD; PRT; 1902 AA.
AC 002470;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE P11-type proteinase precursor (EC 3.4.21.96) (lactocoeprin) (cell wall-
DE associated serine proteinase) (LP151).
GN P11P.
OS Lactobacillus paracasei.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxId=1597;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCDO 151;
RX MEDLINE=92381481; PubMed=1512565;
RA Holck A., Naes H.;
RT "Cloning, sequencing and expression of the gene encoding the cell-
RT envelope-associated proteinase from Lactobacillus paracasei subsp.
RT paracasei NCDO 151."
RL J. Gen. Microbiol. 138:1353-1364(1992).
RN [2]
RP SEQUENCE OF 189-196.
RX MEDLINE=92226694; PubMed=1564442;

```

RA Nies H., Nissen-Meyer J.;
 RT "Purification and N-terminal amino acid sequence determination of the
 RT cell-wall-bound proteinase from *Lactobacillus paracasei* subsp.
 RT *paracasei*."
 RL J. Gen. Microbiol. 138:313-318(1992).
 CC - FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
 CC GROWTH OF THE BACTERIA ON MILK.
 CC - CATALYTIC ACTIVITY: Endopeptidase activity with very broad
 CC specificity, although some substrate preference have been noted,
 CC e.g. large hydrophobic residues in the P1 and P4 positions, and
 CC Pro in the P2 position. Best known for its action on caseins,
 CC although it has been shown to hydrolyse hemoglobin and oxidized
 CC insulin B-chain.
 CC - SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
 CC SUBTILASE FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: M83946; AAA25248.1; -;
 DR PIR: B44858; B44858.
 DR HSSP: P00782; 1S01.
 DR MEROPS: S08.019; -;
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR003137; PA.
 DR InterPro: IPR002029; Peptidase_S8.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR Pfam: PF02225; PA; 1.
 DR Pfam: PF00082; Peptidase_S8; 3.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_SRS; 1.
 DR PROSITE: PS00138; SUBTILASE_HIS; 1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
 DR HydroLase: Serine protease; Cell wall; Zymogen; Signal;
 KW Transmembrane.
 KM
 FT SIGNAL 1 33
 FT PROPEP 34 187
 FT CHAIN 188 1902
 FT DOMAIN 187 1876
 FT TRANSMEM 187 1895
 FT DOMAIN 1896 1902
 FT ACT_SITE 217 217
 FT ACT_SITE 281 281
 FT ACT_SITE 620 620
 FT DOMAIN 1867 1872
 FT SO SEQUENCE 1902 AA; 200253 MW; D8C9F38CE5DA582 CRC64;
 PROTEINS.
 alignment_scores: POTENTIAL.
 Quality: 194.50 Length: 1327
 Ratio: 0.345 Gaps: 65
 Percent Similarity: 42.502 Percent Identity: 19.819
 alignment_block:
 US-09-303-518D-649 x P2P_LACPA ..
 Align seg 1/1 to: P2P_LACPA from: 1 to: 1902
 415 AAAATGTGAAACGGAATTAATAAGACGACCTAAAGCCATCTTCA 464
 |||||
 836 LysIleValaspSerLeuasn.....GlyIleThrTyIserProal 849
 |||||
 465 TGGCGCGCATTTAT...CATATGCCCGCTTG.....C 493
 |||||
 849 agIlygIysnTygIyThValProLeuThrAsnLysAsnThrIyH 866

494 ATAAATTT.....GTCCAGATGCAGACACCTGTGAAATGACC 531
 |||||
 866 IsGIntyTyIyGlyGlyMetValThrAspAla..... 876
 |||||
 532 AGTATATATGATGGCGGGAATATATATGATCAATAATATACCCGACCG 581
 |||||
 877AspGlyLysGIntHValaspAspGlnAla..... 888
 |||||
 562 TGTTCGATTGGGGCAGCAGCAGCAATATTGGCGATCGATGAAGATGAC 631
 |||||
 889PheSerSerAsp..... 892
 |||||
 632 CCAATACCGCGGAAGTTTCATATCATATATGCAAGTCGATTTCTGCTC 681
 |||||
 893LysAsnAlaLeuTyIrsnAspIleSerMetGIntyTyIleu 906
 |||||
 682 GTTGTGGCAATACCTTTCACAA.....AATGATCAGGTG 719
 |||||
 907 LeuATGAsnIleSerAsnValGlnValaspIleLeuAspGlyGlnGlyAs 923
 |||||
 720 TGGCAGACGTCACATAGTAGTGAATAAAATTAACATAGCCCATATGTT 769
 |||||
 923 nlysvallThrThrLeuSerSerSerThrAsnGIntHThrIyThTyI... 938
 |||||
 770 TTTTACCAACAGAGAGCTCATTTGGCGACAGCTGCTCACCAATGTTATC 819
 |||||
 938 938
 |||||
 820 TATGATGCCCAAAAGCAAAAGTGAATTAATGAGGATTTGCAACGGG 869
 |||||
 939 TyIrsnPalHisSerGlnIyIstIyIle..... 947
 |||||
 870 CAACCCCTATATAGAAAAAGCAATGCTTCACGCTGCTGTAAGATT 919
 |||||
 947 947
 |||||
 920 GGTTCATATGTAATCTTGTGGAGATACCCATTCAGTATTCAGCA 969
 |||||
 948 ..TyIrsn.....AlaProIatIrpAspGlyThTyIyIyIasp 960
 |||||
 970 CCAGTCGCAAAATGGAATATCTTTTACAGCAGATAATATGCGACAGG 1019
 |||||
 961 GlnArg...AspGlyLysnIleIyThrAlaaspAspGlySerTyIyThTy 976
 |||||
 1020 AAAATCATGTCCAACATGACACACATCTGCTATATAGATTAATAA 1069
 |||||
 976 rArgIleSerGlyValProGlnGlyIyIyIasp..... 986
 |||||
 1070 CAGCAACCGTTCATATGTTTATGTTTATTCGAGACAGCAGAGAA 1119
 |||||
 987LysArgGlnValPheAspValProPheIyIyIyIyIyIyIyIyIy 1001
 |||||
 1120 CCT...GTTATCATGCTGCGAGGTGCTGTCACACAGTATTCAGCCAGACT 1166
 |||||
 1002 ProThrValArgHisValAlaIleuSerAlaIyIyThr..... 1013
 |||||
 1167 GAATTAATGAGAAATTTCTTTATTCGAGAGGAAAGGCGAATGCA 1216
 |||||
 1014GluAsnGlyIyIyThGlnIyIyT 1021
 |||||
 1217 TACTTACGAGACATCATCATCAAGGTGCTGAGAGATTAATTTCCACAGA 1266
 |||||
 1021 yIleuThrAlaGlnAlaIyIyAspAspIyIyIyIyIyIyIyIyIyIyIy 1033
 |||||
 1267 GATTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1316
 |||||
 1034 AspAlaThrIyIyIySer..... 1038
 |||||
 1317 TCATATCATGTAAGACAGATACCTTACTTGAAGTAACGGCGTGACAA 1366
 |||||
 1039ValIyIyThrAlaIleAsnGlnValIyThr 1048

1367 ACGACCGCTTCCTCAAAATCGCAAGGCGACGCTGCAGCTTCAGCAAA 1416
 1048 snleuaspalathrphethraspalaglythr..... 1058
 1417 GGGGAAAACCAAGGCTCGATCAGCGTGGCGACGCT..... 1452
 1059ThraspalaglythrThrylelel 1067
 1453 .ACAGTCATTTGGATCAGCAGGAGCAATTAAGGCAAAACAGAGCT 1501
 1067 uthrProleuSerAspGlnalaglnalaleucllyasnlglyaspans 1084
 1502 TTGAGTAATCGGCTTGGTCACGCGCAGGGGTACGGTCAACTGAATGCC 1551
 1084 erlaleucllythrleuThrAsp.....Asnala 1093
 1552 GATTAATCACTTCACCCGACAAACTCATTTTC.....GGCTTCCGCG 1595
 1094 SerasnalaThrAsnclnaspalaservalelnlySproglySerThrse 1110
 1596 CGGAGCTTTGGATTTAAACGGGCAATTCGCTTCCGTCACCGATTCGAAA 1645
 1110 rPheaspLeuileValasnlglyglyllePro...AsplysleSers 1126
 1646 ATACCGATGAGGGCGGATGTGTCAACCAATCAAGCAAAAGATTC 1695
 1126 erThrThrThrly.....TyrluinalasnThrclnglyglylle 1139
 1696 ACCGTTACCATTTACAGCAATTAAGATTTGTCTACAACGGCAATTAACA 1745
 1140 ThrlyThrPheserGlyThrTyProialaleValaspGlyThrlyThr 1156
 1746 CAGCTTGATAGCAAAAGAAATTCGCTTCAACGGTGGTGGGAGA 1795
 1156 rAspalaglncllylyshis..... 1163
 1796 AAGATACGACCAAAACGACGGGCGCTCACTTGTTCACGCGCGCC 1845
 1164AspleuasnThrThrlyraspalaala 1172
 1846 GCAGAAAGCCGACCCCTGCTGCTT.....TCGGG 1874
 1173 ThrAsnSerPhethrAlaSerMetalaValThrAsnAlaaspryAlaAl 1189
 1875 CGAACAATAATTAAACGGCAATCAGCAAAACGCAAAACGCTTT 1924
 1189 aglnValaspLeuTyAlaaspLyAlaHisThrclnleucllyshisp 1206
 1925 TCAGCGCGACCA.....ACACGCGACGCTTAC 1953
 1206 heaspThrlyValargleuthrAlaProthrPhethrAspleuLySphe 1222
 1954 AATTCATTTAAACGACCATTTGTCGCAAAAGAGGCAATTCCTCCGGGGA 2003
 1223 AsnansglySerAspGlnThrSerclnularThrlylelyValThrGlyTh 1239
 2004 AATCGTGGGACACGACGATCAGCAACCGCAATTTAAAGCGCAAACT 2053
 1239 rValSerSerAspThrlyThrValAsn..... 1248
 2054 TCCAATTAAGGCGACGCGCTGCTTCCGCAATTTGCCAAAGTG 2103
 1249ValGlyAspThrValAlaAlaLeu 1256
 2104 AAGGCGATTTGGATTTGAGCAATCAGCCCAAGCAATTTTGGTGGC 2153
 1257 AspalaglnHisPheserValaspValProvalasnTyrcly..... 1271
 2154 ACCGATCAACGACCAATC...TGTACAGCTTCGAGTGCAGCGGTC 2200
 1272AspasnThrlylelyValThrAlaThrAspGlnaspGlyAla 1285
 2201 TGACAATTTGTGCGAAAAAACCAATTAC.....GACGATAAAGTG 2241

1285 snThrThrThrGlnGlnlyThrThrThrSerSerTyAspProaspVal 1301
 2242 ATTGCTTCAATGACTAAGACGAC.....ATCAGCGCAATGT 2279
 1302 LeuLyAsnAlaValThrPheaspGlnGlyVallySphelyAlaAsnG 1318
 2280 CGATCTTCGATTCACGCTCATTTA.....AATTCACAGGGCTGCCA 2323
 1318 urPheasnAlaThrSerAlaLySphetyrAspProlyThrcllylelAlat 1335
 2324 CACTACAGGCG.....AATTCAGTGCATATGCC 2352
 1335 hrleThrcllylyVallyshisProthrThrThrleuGlnValaspGly 1351
 2353 GATTAACGCTTATACGTAGCAGCAGCAGCCCAACGCAACCTTAG 2402
 1352 ...LysGlnleSerlelyAsnaspLeuThrPheSerPhethrleuS 1367
 2403 CCTC.....GTGGCAATGCC. 2418
 1367 pleucllyThrleucllyGlnlySprophelyValValGlyAspThr 1384
 2419CAAGCAACATTTAATCAAGC..... 2439
 1384 hrGlnasnlyThrPheGlnGlnAlaThrPhethrleuSaspAlaVal 1400
 2440ACATTAACGGCAACATTCGCT.....TCGGG 2468
 1401 AlaprothrleuSerleuAspSerThrAspAlaProvalTyThrAs 1417
 2469 CAATGCTTCAATTAATCAAGCAGCAGCCGCTCAAAACGCAAGCTGA 2518
 1417 nasPProasnPhelGlnleThrcllyThrAlarhAspans..... 1430
 2519 CGCTTCCGCAACGCTAAGCAACGTAAGCATTCGCACTCAAGCT 2568
 1431AlaGlnTyLeuSerleuAlaIleasnGly 1440
 2569 AATGCTCCCTACCGATAGGACAGTATTCATTTGAACGACGCGCTT 2618
 1441 Ser.....HisValAlaSerGlnTyAl 1448
 2619 TACCGACAATCAGCGGCGCAAGGATACGCAATTAAGACA 2668
 1448 asplyleasnIleasnSerGlylySpro..... 1457
 2669 GCGAATGAGCGTGGCTGACGACGCAATTAAGCAATTTAAGCTTAC 2718
 1458GlyHisMetAlaIleasp 1463
 2719 AACGCGACCATTTACATTCATTCGCGCTATTCGCAAGATGGCGAGGCG 2768
 1464 GlnProvallySleuLeu.....GlnGlylySasnVa 1474
 2769 GCAAAACGCGAGTGAGCAGATGCGCGCGCGCTTCCGCGCTTGC. 2817
 1474 lThrValAlaValThrAspSerGlnAsnAspThrThrlylySyl 1491
 2818GCGCTTCCCTAATTA.....TCGTTTACA 2841
 1491 leThrValTyTyrgluProlylySphThrleuAlaAlaProthrValThr 1507
 2842 CGGCAACCTTCGATAGATCCGCTTCAACAGCAGTACGCTTAACGCA 2891
 1508 ProserThrThrGluProAlaLySphValThrleuThrAlaAsnAla 1524
 2892 ATTGAAGGCGAC..... 2904
 1524 alaThrcllyGlnThrValGlnTySerAlaaspGlylySphThrTyG 1541
 2905GGAACATTTCCGC 2916

1541 InAspValProAlaIaGlyValThrValThrAlaAsnGlyThrPheLys 1557
2917 TTTATGCG...GAACCTTCGGCTACCGACGCAAAATTAAGCTGCG 2963
1558 PheLysSerThrAspLeuTyrGlyAsnGlySerProAlaValAspTyr.. 1573
2964 GGAAGTTCGGAAGGCACTTACACCTTGGGGTCAACATTAACGGCAGC 3013
1574ValValThrAsnIleLysAlaAspA 1582
3014 AACCTGCAGCGCTGACAAATTAAGCTAGTGGAGAAAACAAACAA 3063
1582 sProAla.....GlnLeuGlnThrAla LysGlnAlaLeuThrAs 1595
3064 CCGGTGTCGGAACCTTAATTTACCCCTGCAAAAGCAACGTCGATGC 3113
1595 n.....LeuIleAlaSerAlaLysThr.....LeuS 1604
3114 CGGGCGGTGGGTTACCACTCATCCGCAAAAGCGCGAGTTCGGCTGC 3163
1604 erAlaSerGlyLysTyrAspAspAlaThrThrAlaLeuAlaAla... 1619
3164 ATATCCGGTCAAGACAGAGCTTTCGCAACACTCGCAAGGAGAA 3213
1620AlaThrGlnLys 1623
3214 GCCAAAAACAGCGCGGAAAAAGACACCGCAAGCCTTGACGCGCTGAT 3263
1623 sAlaGlnThrAlaLeuAspGlnThrAsp..AlaSerValAspSerLeuTh 1639
3264 TGGCGCGCGGCGCGATCCGTCGAAAGACAGAAAGCCTGCCGACCGG 3313
1639 rGlyAlaAsnArgAspLeuGlnThrAlaIleAsnGlnLeuAla..... 1653
3314 CCGCGAGCGAGCGGGGAAAAATGTCGATATGACAGCGAGAGAGAG 3363
1653 1653
3364 AAAAAGGGGTGCGAGCGGATAAAGACACCGCTTGGCAACACGCGA 3413
1654 ...AlaLysLeuProAlaAspLysLysThrSerLeuLeuAsnGlnLeuG 1669
3414 AGCG.....GAACCGCGCGG 3430
1669 nSerValLysAlaAlaLeuGlyThrAspLeuGlyAsnGlnThrAspPro 1686
3431 CTACC...ACGCGCTTCCCGCGCGCGCGCGCGCGCGGATTTGGCG 3477
1686 erThrGlyLysThrPheThrAlaAla.....LeuAsp 1696
3478 CAACCTGCACCCCAACCGCAGCCCAACCGCAGCGGACCTGATCAGCG 3527
1697 AspLeuValAlaGlnAlaGlnAlaGlyThrGlnThr..... 1708
3528 TTAATGCCAATAGCGTTGAGTGAATTTCCGACGCTCAACAGCGTTT 3577
1709AlaAspGlnLeuGlnAlaSerLeuAlaLysVal.. 1719
3578 TCGCGGTACAGAGCAATTAAGCGCGCTATTTGGCGAG..... 3616
1720LeuAspAlaValLeuAla..LysLeuAlaGlnG 1730
3617 ...ACGGCGGACGCGG.....TTTGACACAGCGGAT 3647
1730 yLleLysAlaAlaThrProAlaGlnValGlyAsnAlaLysAspAlaAla 1747
3648 CCGGG.....ACACCAACACTACCGTTGCGAAG 3676
1747 hrGlyLysThrTyrPyrAlaAspIleAlaAspThrLeuThrSerGlyGln 1763
3677 ATTTCGGCGCGCTACGCGCAACCAACGACCTGCGCAATGATGCGAG 3726
1764 AlaSerAlaAspAlaSerAspLysLeuAlaHisLeuGlnAlaLeuGlnSe 1780

3727 AAAAACCCTGCGACGCGGCGCTCGGCAATCCGTTTTCGCAACACGAGC 3776
1780 rLeuLysThrLysValAlaAlaValaGlnAlaAlaLysThrAlaGly.. 1796
3777 CGAAACACCTTCGACGAGCGCATCGGCAACTCGGCACGCG..... 3817
1797 ..LysGlyAspAspThrThrGlyThrSerAspLysGlyGlyGlnGly 1812
3817 3817
1813 ThrProAlaProAlaProGlyAspThrGlyLysAspLysGlyAspGlnG 1829
3818TTGCCACAGCGCGCGCTTTTGGGCGCA 3842
1829 ySerGlnProSerSerGlyLysAsnIleProThrLysProAlaThrThrT 1846
3843 ATACGCGATCGACAGGTTCATCATCG 3868
1846 hrSerThrSerThrAspAspThrThr 1854